AD-A267 514

ARI Research Note 93-18

Preliminary Validation of the Task Analysis/Workload Methodology

David B. Hamilton and Kenneth D. Cross
Anacapa Sciences, Inc.

ARI Aviation R&D Activity at Fort Rucker, Alabama
Charles A. Gainer, Chief

Training Systems Research Division
Jack H. Hiller, Director

June 1993

United States Army
Research Institute for the Behavioral and Social Sciences

U.S. ARMY RESEARCH INSTITUTE FOR THE BEHAVIORAL AND SOCIAL SCIENCES

A Field Operating Agency Under the Jurisdiction of the Deputy Chief of Staff for Personnel

EDGAR M. JOHNSON
Acting Director

Research accomplished under contract for the Department of the Army

Anacapa Sciences, Inc.

Technical review by

John E. Stewart
Dennis C. Wightman

Accesion For	
NTIS CRA&I	☑
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution /	
Availability Codes	
Dist	Avail and / or Special
A-1	

DTIC QUALITY INSPECTED 3

NOTICES

DISTRIBUTION: This report has been cleared for release to the Defense Technical Information Center (DTIC) to comply with regulatory requirements. It has been given no primary distribution other than to DTIC and will be available only through DTIC or the National Technical Information Service (NTIS).

FINAL DISPOSITION: This report must be destroyed by any method that will prevent disclosure of contents or reconstruction of the document. Please do not return it to the U.S. Army Research Institute for the Behavioral and Social Sciences.

NOTE: The views, opinions, and findings in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other authorized documents.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	June 1993	Interim Jul 90 - Dec 91

4. TITLE AND SUBTITLE
Preliminary Validation of the Task Analysis/Workload Methodology

5. FUNDING NUMBERS
C-MDA903-87-C-0523
63007A
795
3405
C06

6. AUTHOR(S)
Hamilton, David B.; and Cross, Kenneth D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Anacapa Sciences, Inc.
P.O. Box 489
Fort Rucker, AL 36362-5000

8. PERFORMING ORGANIZATION REPORT NUMBER
ASI690-353-92

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Research Institute for the Behavioral and Social Sciences
ATTN: PERI-IR
5001 Eisenhower Avenue
Alexandria, VA 22333-5600

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
ARI Research Note 93-18

11. SUPPLEMENTARY NOTES
Contracting Officer's Representative, Charles A. Gainer

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release;
distribution is unlimited.

12b. DISTRIBUTION CODE
--

13. ABSTRACT (Maximum 200 words)

Over the past 8 years, the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI) has sponsored the development of the Task Analysis/Workload (TAWL) methodology. The methodology is used to develop computer-based models that predict the operator workload of Army weapon systems. The TAWL methodology has been used to predict operator workload in existing and modified versions of the AH-64, UH-60, and CH-47 aircraft. This preliminary research was designed to assess the validity of the methodology. Seven experimental task conditions, analogous to rotorcraft operation and designed to vary in workload, were used in the research. Two TAWL workload prediction models of the conditions were independently developed by two analysts. Twenty AH-64A aviators performed repeated trials in each of the task conditions. Measures of the aviators' task performance, subjective workload, and physiological workload were obtained as TAWL model validity criteria. Although there were some differences between the models developed by the two analysts, the predictions generated by the models were highly correlated. In addition, the correlations between the average TAWL predictions for both models and the criterion measures were

(Continued)

14. SUBJECT TERMS
Workload
Aviator workload
Function analysis
Workload prediction
NASA-TLX
Workload measures
Aviator performance
Task analysis
Modeling
Mission analysis

15. NUMBER OF PAGES
147

16. PRICE CODE
--

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

NSN 7540-01-280-5500

Standard Form 298 (Rev 2-89)
Prescribed by ANSI Std Z39-18
298-102

13. ABSTRACT (Continued)

significant and high (between .89 and .99). The results support the validity of the TAWL methodology.

FOREWORD

The U.S. Army Research Institute Aviation Research and Development Activity (ARIARDA) at Fort Rucker, Alabama, is an operational unit of the U.S. Army Research Institute for the Behavioral and Social Sciences (ARI). This work was performed as part of the front-end analysis work within ARIARDA's research mission. The work was performed under the Memorandum of Understanding between the U.S. Army Aviation System Command (AVSCOM) and ARI, "Establishment of Technical Coordination Between ARI and AVSCOM," 10 April 1985.

The potential impact of advanced technology on manpower and personnel requirements must be considered when managing aviation systems. Since high operator workload can result in a dramatic decrease in system effectiveness, workload is a critical consideration.

Over the past 8 years, ARIARDA has produced microcomputer-based models of operator workload for existing and modified versions of several series of aircraft. The methodology used to develop the models has produced organized and useful results; however, it has not been validated until now. This report describes research designed to test the validity of the methodology.

The results of the research have been briefed to ARI, AVSCOM personnel, and U.S. Army Aviation Center directorates. The validated workload prediction methodology should continue to be useful in evaluating the operator workload in Army systems well into the future.

ACKNOWLEDGMENTS

The authors wish to express their appreciation to all of those who contributed to the project. From the U.S. Army Research Institute Aviation Research and Development Activity, Charles Gainer provided the overall direction for the research and Major Dale Weiler coordinated research subjects.

From the Human Engineering Group of the Armstrong Aerospace Medical Research Laboratory at Wright-Patterson Air Force Base, Ohio, we would especially like to thank Major Albert Badeau for his patient instruction in the use of the heart interbeat interval analysis software and for conducting the frequency-based analyses of heart rate variability.

From the Anacapa Sciences staff, thanks to Monty Grubb, Gary Coker, Steve Lewis, Ken Persin, and D. Michael McAnulty for their assistance in executing the research and preparing the report.

PRELIMINARY VALIDATION OF THE TASK ANALYSIS/WORKLOAD METHODOLOGY

EXECUTIVE SUMMARY

Requirement:

The research described in this report was conducted to provide a preliminary indication of the validity of the Task Analysis/Workload (TAWL) methodology. Other objectives were

- to determine the differences in TAWL models produced independently by two qualified analysts and to assess the effect of the differences on TAWL predictions and validity,
- to evaluate the mechanisms used in the methodology that link TAWL predictions to operator performance, and
- to determine the relationship between operator heart rate variability and (a) objective performance measures, (b) subjective ratings of workload, and (c) TAWL predictions.

Procedure:

Because a flight or simulation environment was unavailable, a set of conditions was developed that produced a large range of operator workload. The equipment used to generate the conditions included a computer and display, a voice synthesizer, a numeric keypad, a foot controller, and two hand controllers. A dichotic listening task, a dual-axis tracking task, and two independent single-axis tracking tasks were used singly and in combination to produce seven different conditions that varied in workload. Two analysts independently developed TAWL workload prediction models of the seven conditions. Twenty male AH-64A aviators (10 battle-rostered crews) from one battalion of an operational aviation regiment performed repeated trials of the task conditions over three test sessions. Subjective workload ratings, measures of tracking and dichotic listening performance, and measures of aviator heart rate variability were taken for each trial.

Findings:

The conduct of the research resulted in the following individual findings:

- Seven conditions that vary in workload can be constructed from combinations of four tracking and dichotic listening tasks.
- Two qualified analysts were able to use the TAWL methodology to construct workload prediction models and generate predictions of the workload in each of the conditions.
- Three types of differences (structural, rating, and opinion) were found in the construction of the two models.
- The models' predictions differed but were highly correlated ($r = .99$).
- The subjective workload ratings were well organized and sensitive to the task loading differences between the conditions.
- The tracking performance consistently increased as a function of condition and had reached a plateau by the last practice session.
- Two measures of dichotic listening performance were well organized and sensitive to task condition.
- The heart rate variability measures were not well organized or sensitive to task condition.
- Three measures showed sufficient sensitivity to the across-condition workload manipulation to be used as criteria for TAWL validity.
- The correlations among the three criterion measures and the average of TAWL component predictions were high ($.89 < r < .99$) for both models.

These findings led to the following conclusions:

- The TAWL methodology has excellent potential for generating valid predictions of operator workload.
- Although differences were observed between the models generated by different analysts, the differences did not reduce the validity of either model.

- Although the relationship between operator workload and performance is complex (even small amounts of workload can sometimes degrade performance), the TAWL methodology provides an excellent description of the aspects of the task environment known to affect operator performance.
- The validation of the methodology should continue with the validation of a TAWL model of full complexity.

Utilization of Findings:

The findings support the validity of the TAWL methodology for producing valid models of operator workload. Operator workload should continue to be a critical consideration in Army system design well into the future.

PRELIMINARY VALIDATION OF THE TASK ANALYSIS/WORKLOAD METHODOLOGY

CONTENTS

CONTENTS (Continued)

LIST OF TABLES

Page

LIST OF FIGURES

GLOSSARY OF ACRONYMS AND ABBREVIATIONS

AQC - aviator qualification course
ANOVA - analysis of variance
ARI - U.S. Army Research Institute for the Behavioral and Social Sciences
ARIARDA - ARI Aviation Research and Development Activity
AUD - auditory
AVSCOM - Aviation System Command
COG - cognitive workload
CMS - combat mission simulator
DL - dichotic listening
DRA - data recording and analysis
EF - effort
FR - frustration
IBI - interbeat interval
IERW - initial entry rotary wing
ITEMS - interactive tactical environment management system
LHX - light helicopter, experimental
KIN - kinesthetic
MD - mental demand
MSW - mean subscale workload
NASA - National Aeronautics and Space Administration
OC - overload conditions
OD - overload density
OP - own performance
OW - overall workload
PD - physical demand
PSY - psychomotor
PW - peak workload
SME - subject matter expert
STRATA - simulator training research advanced testbed for aviation
TAWL - task analysis/workload
TD - temporal demand
TLX - task load index
TOSS - TAWL operator simulation system
USAAVNC - U.S. Army Aviation Center
VIS - visual
WAM - workload assessment monitor

PRELIMINARY VALIDATION OF THE TASK ANALYSIS/WORKLOAD METHODOLOGY

Introduction

A methodology was developed for predicting operator workload using the information from a task analysis of the system. The original methodology was developed for use in the concept exploration and definition phase of the Army's light helicopter, experimental (LHX) development program (Aldrich, Craddock, & McCracken, 1984; Aldrich, Szabo, & Craddock, 1986; McCracken & Aldrich, 1984). Analyses were conducted to compare the operator workload of one- and two-crewmember configurations of the LHX.

Subsequently, a comprehensive task/workload analysis of all phases was conducted for the AH-64A attack mission. The results of the analysis are described in a technical report by Szabo and Bierbaum (1986). Applying the methodology to an operational aircraft (AH-64A) allowed a far greater degree of specificity in the task element descriptions than was possible during the concept exploration and definition phase of the LHX program.

The methodology used to perform the AH-64A analysis was further refined and used to predict the effect on operator workload of modifications to Army special operations helicopters. The methodology was used to predict the crewmember workload for existing and modified versions of the UH-60A (Bierbaum & Hamilton, 1990; Bierbaum, Szabo, & Aldrich, 1989), and CH-47D aircraft (Bierbaum & Aldrich, 1989; Bierbaum & Hamilton, 1991).

The refined version of the methodology is called the task analysis/workload (TAWL) methodology. Computer support for the methodology has been developed and named the TAWL operator simulation system (TOSS). Hamilton, Bierbaum, and Fulford (1991) provide a detailed description of both the TAWL methodology and the TOSS software.

The TAWL Methodology Overview

A TAWL workload prediction model is developed in three stages. In the first stage, the analyst performs a task/workload analysis on the system. A prototype mission for the system is developed and is progressively decomposed into phases, segments, functions, and tasks. The analysis yields estimates of the duration of tasks, a description of the sequence of tasks, and a description of the crewmember and subsystem associated with each task. The workload analysis, based on a multiple resources theory of human attention, yields independent estimates of the cognitive, psychomotor, and sensory components of workload (hereafter referred to as workload components) for each task. The theory differs from other multiple resource theories of attention in the nature and number of components that are identified. The theory underlying the TAWL methodology recognizes five independent workload components: auditory, kinesthetic, visual, cognitive, and psychomotor. See Wickens (1984) for a review of multiple resource theories of attention and their relation to workload.

The TAWL methodology treats each of the workload components independently for two reasons. First, although interactions among the components probably occur, the nature of the interactions cannot be defined adequately at this time. Second, the information that results from treating workload components individually is useful for identifying potentially effective ways to reduce workload or to redistribute workload among the crewmembers, subsystems, or components. For example, a designer can decide whether additional information should be presented visually or aurally by determining which component has the least amount of workload.

The workload analysis is based on subjective estimates of operator workload rather than on estimates derived through experimentation. Research analysts and subject matter experts (SMEs) generate workload estimates by using equal-interval, verbally anchored rating scales; the scale values range from 0 to 7. This approach avoids the large expenditures of time, money, and manpower that would be required to derive empirical measures of workload for each task.

In the second stage of the TAWL methodology, the analyst develops a model of each crewmember's actions by recombining tasks to simulate the behavior of the crewmembers during each segment of the mission. Function decision rules are developed that describe the sequencing of tasks within each function; segment decision rules are developed that describe the start time, stop time, and interaction of the functions within each segment. The underlying assumption is that the segments can be combined to model the crewmembers' behavior for individual mission phases and for the entire mission.

In the third stage of the TAWL methodology, the analyst executes the model to simulate the crewmembers' actions during the operation of the system. The TOSS computer software performs the simulation and produces estimates of each crewmember's cognitive, psychomotor, and sensory workload for each half-second of the mission. The estimates of workload for each component are generated by summing the workload for that component across all tasks that the crewmember performs during each half-second of the mission. For example, during a specific half-second interval, the pilot performs the following tasks: Control Attitude, Check External Scene, and Transmit Communication. The cognitive workload for the three tasks during that interval is 1.0, 1.0, and 5.3, respectively. Thus, the estimate of cognitive workload for the pilot during that interval is 7.3. A predefined overload threshold is used during execution of the model to measure the amount of time during the mission that each crewmember experiences an overload condition.

Using the TAWL prediction methodology, an analyst can develop a model of a system and use the model's output to determine:

- the absolute and relative workload of the crewmember,
- the time intervals during which crewmembers experience high workload, and
- the components for which crewmembers experience high workload.

The TAWL methodology yields sufficient information to enable system designers to reduce or redistribute workload over time, crewmembers, or components. Designers also may use the information to identify design alternatives that result in lower workload. In addition to the uses described above, the methodology yields mission timelines and task listings that can be used to develop the system's manning and training requirements.

TAWL Validity Research

The only research conducted to test the validity of the predictions generated by a TAWL workload prediction model was performed by Iavecchia, Linton, Bittner, and Byers (1989). The researchers collected subjective ratings of overall workload (OW) and peak workload (PW) from seven UH-60A crews for six segments of a mission flown in the UH-60A flight simulator. They used the UH-60A workload prediction model (Bierbaum & Hamilton, 1990; Bierbaum, Szabo, & Aldrich, 1989) to compute OW-comparable TAWL predictions by averaging across the six workload components (i.e., cognitive, psychomotor, auditory, visual-unaided, visual-aided, and kinesthetic) and each half-second of segment time. They prepared PW-comparable TAWL predictions by summing across workload components for each half-second of segment time and by selecting the maximum sum for each mission segment.

The TAWL predictions correlated highly with OW for the 12 unique conditions (r =.82). The researchers noted that the TAWL predictions were consistent with all of the copilot OW ratings and all but one of the pilot ratings. The pilot communication in this segment was not as complex as assumed in the model. Subsequently, this segment was dropped from the analysis and the correlation was recomputed using the remaining 11 segments; the correlation coefficient increased to .95. The correlations between TAWL and PW were lower, r = .62. Based upon these results, Iavecchia et al. (1989) concluded, "The TAWL model has substantial potential as an analytical workload estimation technique which may be applied before system development."

Research Objective

Although the results of Iavecchia's research (Iavecchia et al., 1989) support the validity of TAWL predictions, the research has four important limitations. First, the research employed only subjective workload ratings as validation criteria. Second, the subjective workload ratings used as validation criteria (OW and PW) are based on each aviator's own internal (unknown) reference and not on a standard reference task. Third, the relationship among TAWL predictions and subjective workload ratings was determined for a limited range of task loadings. Finally, a relatively small number of different conditions was examined in the research. The present research addresses the first three of the limitations listed above, but not the fourth. Specifically, the objective of the present research was to assess the validity of TAWL predictions for a range of task loading using both objective performance measures and a well established subjective rating of workload--the NASA Task Load Index (TLX)--as validation criteria.

The present research had three additional objectives. First, determine the differences in TAWL models produced independently by two qualified analysts and assess the effect of the differences on TAWL predictions and validity. Second, evaluate the mechanisms used in the methodology that link TAWL predictions to operator performance based on the results of the research and refine the methodology. Third, determine the relationship between operator heart rate variability and (a) objective performance measures, (b) subjective ratings of workload (TLX), and (c) TAWL predictions. Because heart rate variability has been proposed as a valid and sensitive measure of workload (Aasman, Mulder, & Mulder, 1987), the present research was used as an opportunity to explore the ease of use and sensitivity of the measure in the aviation task environment. If heart rate variability proved to be a sensitive and viable measure of the task load manipulations of interest, it could be used in future research to provide a continuous, non-intrusive measure of workload that could be acquired in flight.

Method

This section describes the methods used to assess the validity of the TAWL workload prediction methodology. The following eight subsections describe the experimental tasks, model development, TAWL predictions, materials, aviators, apparatus, procedures, and data analysis.

Experimental Tasks

Task selection. To examine the relationship among the variables of interest over a wide range of conditions, a set of tasks was needed that produces a large range of operator workload. Fortunately, the U. S. Army Aviation Center (USAAVNC) uses just such a set of tasks to aid in the assignment of aviators to aircraft after they have completed initial entry rotary wing (IERW) training (Intano, Howse, & Lofaro, 1991a; Intano, Howse, & Lofaro, 1991b; Intano & Lofaro, 1989; Intano & Lofaro, 1990). The tasks were developed from research conducted by the Israeli Air Force (Gopher & Kahneman; 1971) and were adapted by the Navy for use in aviator selection (Griffin & MacBride; 1986). The tasks are controlled by a computer program written in the Basic computer language. The equipment used to generate the tasks included a computer, a computer display, a voice synthesizer, a numeric keypad, a foot controller, and two hand controllers. Different conditions are produced by using four basic tasks singly and in combination. The four tasks are a dichotic listening task, a dual-axis tracking task, and two independent single-axis tracking tasks.

Dichotic listening tasks. Seven conditions were constructed from combinations of the four basic tasks. The four basic tasks and the seven conditions are described below.

In the dichotic listening (DL) task, standard audio headphones were used to present simultaneously different strings of letters and digits to each ear. The aviators were instructed to attend only to the information presented to one ear and to enter on a numeric keypad the digits in that string. Each trial lasted 3 minutes. During each trial, 8 DL presentations were made using 16 22-character strings for each ear. Each DL presentation consisted of the 27 events listed in Table 1. A trial consisted of an initial period of silence, 8 stimulus presentations, and another period of silence at the end.

To remain consistent with the other DL tasks being used by the USAAVNC, the stimulus strings were presented and analyzed in two parts. The first part (16 characters) consisted of randomly dispersed digits and letters; the second part (6 characters) always consisted of 2 letters followed by 4 digits. This pattern was obvious to the participants and was used in their task performance strategies.

For each DL trial, a random number from 1 to 36 was selected that defined the starting point in the 36 record DL stimulus file. The 36 DL stimulus strings used in the experiment are presented in Table 2.

Tracking tasks. The remaining three basic tasks were continuous, compensatory tracking tasks that required the aviators to use a control to keep a moving cursor aligned with a stationary target. The three tracking tasks differed in the number of dimensions in which the cursor could move (one or two dimensions), the direction of the cursor movement (horizontal, vertical, or both), and the design of the control (single axis joystick, dual axis joystick, or pedals). The target and cursor for all three tracking tasks were presented on the same color computer display. The stimulus configuration for the three tracking tasks is illustrated in Figure 1 and described below.

The target and cursor for the dual-axis tracking task appeared within a large rectangle located in the upper left portion of the display. The target consisted of a large cross created by a vertical dashed line and a horizontal dashed line that intersect at the center of the rectangle; the cursor was a small cross created with solid lines (see Figure 1). The cursor was controlled with a dual-axis joystick. The joystick was mounted between the aviator's legs in a manner similar to the control stick of a fixed-wing aircraft. The pivot point of the dual-axis joystick was located only a few inches below the thigh level. The dual-

Table 1

Timing Sequence of a Single Dichotic Listening Presentation

Sequence	Time (sec)	Event
1	.70	"TEST"
2	1.39	"LEFT" or "RIGHT" (instruction for which ear to attend)
3	2.09	1st character in part 1 (stop collecting responses for part 2)
4	2.79	2nd character in part 1
5	3.48	3rd character in part 1
6	4.18	4th character in part 1
7	4.88	5th character in part 1
8	5.58	6th character in part 1
9	6.27	7th character in part 1
10	6.97	8th character in part 1
11	7.67	9th character in part 1
12	8.36	10th character in part 1
13	9.06	11th character in part 1
14	9.76	12th character in part 1
15	10.45	13th character in part 1
16	11.15	14th character in part 1
17	11.85	15th character in part 1
18	12.55	16th character in part 1
19	13.24	silence
20	13.94	silence
21	14.64	1st character in part 2 (stop collecting responses for part 1)
22	15.33	2nd character in part 2
23	16.03	3rd character in part 2
24	16.73	4th character in part 2
25	17.42	5th character in part 2
26	18.12	6th character in part 2
27	20.52	silence

Note. Each event from sequence 1 to 26 used 23 ticks of the computer clock (≈.7 s) except for action number 27 which used 79 ticks (≈2.4 s). Time is approximate elapsed time in seconds.

Table 2

The 36 Stimulus Strings Used in the Dichotic Listening Task

	Left Ear		Right Ear	
Stimulus	Part 1	Part 2	Part 1	Part 2
1	GZY3FMF49GOSLØ1F	BZ2741	B8FXOF2LS7GR6XY5	FY963Ø
2	LO9NYBN17FXFS35R	GX6381	M8ZGXFØFO2SG6MF4	LS729Ø
3	2XN9FFØMY6SL3OBG	MR9438	S48RBXY75NFGLSF1	NO6152
4	M43LFXN6SY7GO2LF	ZB2897	9ZY8GF1ORØSF5RNX	YF6413
5	Z6YO3FLZØXM9NL7F	XG2147	2LXNMYF4S51RGBR8	SLØ936
6	YF7FLRG13ONOR94G	RM4126	X2NSYM8LF6BXØFY5	ON5398
7	GFSF1NY85RNØFR9Z	BZØ927	LY7GX63LF2SMOXN4	FY1836
8	L4SGØFZX6NO2XFM8	GX2516	5ROFR19NS37LYBLO	LS8349
9	F1XNO78RLY5MFFS4	MR3146	BGSLØML93OY6XN2Z	NO7982
1Ø	F5FO8LMSØFG6YMX2	RB639Ø	4GNFO17NL93OGRYF	SF7421
11	L81RO4XNGBS5MX2L	LM8935	7YF9LNYOMLØF3RZ6	XO6214
12	O5LR2OFX6OS7GRB8	GRØ369	1OMSF43YBØ9GFXFZ	FN1472
13	1FLØOS9GF4FMY3GZ	OG4673	Y56XGRS72LOFFXB8	SL21Ø9
14	5RS3XF7FN1YB9NLO	NF9753	R46MSGO2ØFXFZGM8	OG6418
15	BG3OSLY6ØMFFN92X	YM4135	F1LXFG5NY7BS8RS4	NB6287
16	LFO27GSYN6FX3LM4	SN1862	NX5RSFRØ1OGFY89Z	RO4Ø95
17	7FNLM9ØXLZ3FYOZ6	XL46Ø9	R8GB1RS5F4MYXN2L	FR5827
18	4GR9NO3OG1LR7FYF	XL1937	Y5ØFBSF68LYMNSX2	BM2845
19	9ZFRNØ5RY81NSFGF	RY4681	N4OXSMF23LX67GLX	LS7935
2Ø	M8XFO26NZXØFSGL4	NZ2768	LOYB7LS39NR1OF5R	RO1453
21	S4FF5MLY8RO7XNF1	RMØ459	2ZXNY63OL9ØMSLBG	LO8126
22	X2YMG6ØFMS8LFOF5	GS7285	YFGR3OL97NO1NF4G	XY9Ø64
23	2LMXS5BGXNO41RL8	SB6485	Z63RØFMLYOLNF97Y	MR7391
24	B8GRS76OFX2OLRO5	XF219Ø	FZFX9GBØ3YF4MS1O	YB6473
25	B8FXOF2LS7GR6XY5	FY963Ø	GZY3FMF49GOSLØ1F	BZ2741
26	M8ZGXFØFO2SG6MF4	LS729Ø	LO9NYBN17FXFS35R	GX6381
27	S48RBFY75NFGLOF1	NO6152	2XN9FSØMY6SL3ZBG	MR9438
28	9ZY8GF1ORØSF5RNF	YF6413	M43LFXN6SY7GO2LS	ZB2897
29	2LXNMYF4S51RGBR8	SLØ936	Z6YO3FLZØXM9NL7F	XG2147
3Ø	X2NSYM8LF6BOØFY5	ON5398	YF7FLRG13ONXR94G	RM4126
31	LF7GX63LF2SMOXN4	FY1836	GFSF1NY85RNØFR9Z	BZØ927
32	5ROFR19NS37LYBLO	LS8349	L4SGØFZX6NO2XFM8	GX2516
33	BGSLØML93OY6XN2Z	NO7982	F1XNO78RLY5MFFS4	MR3146
34	4GNFO17NL93OGRYF	SF7421	F5FO8LMSØFG6YMX2	RB639Ø
35	7YF9LNYOMLØF3RZ6	XO6214	L81RO4XNGBS5MX2L	LM8935
36	1OMSF43YBØ9GFXFZ	FN1472	O5LR2OFX6OS7GRB8	GRØ369

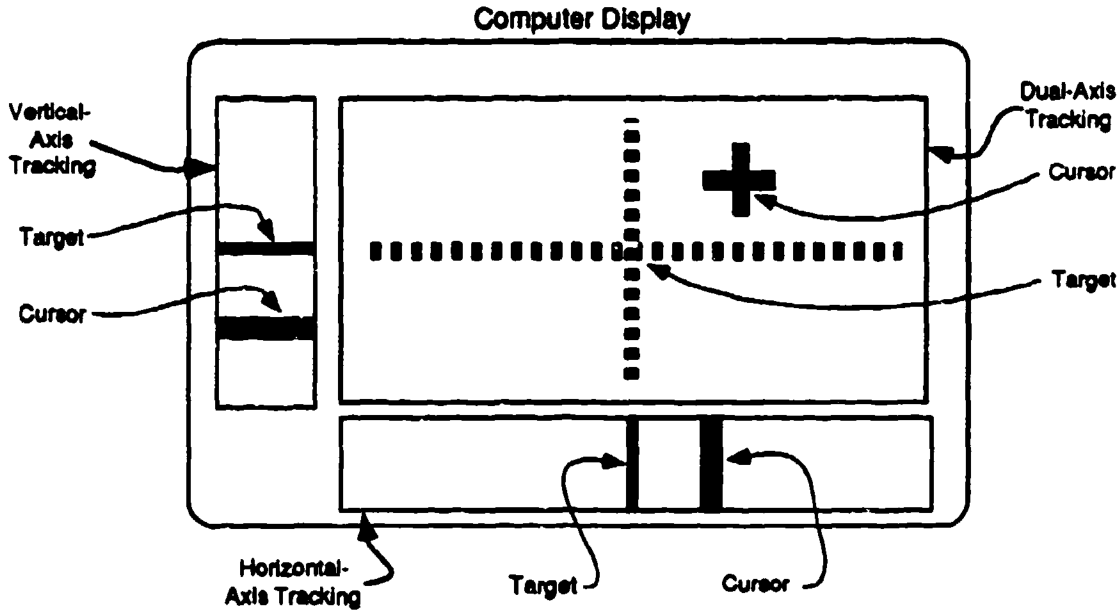

Figure 1. Tracking task display configuration. (The annotation, added here for descriptive purposes, was not present on display).

axis joystick moved in the same planes as an aircraft control stick (right-left and fore-aft). Movement of the joystick in the right-hand (left-hand) direction caused the cursor to move left (right); movement of the joystick in the forward (backward) direction caused the cursor to move down (up).

The target and cursor for the horizontal-axis tracking task appeared in a rectangle located on the lower portion of the display, directly beneath the large rectangle in which the dual-axis target and cursor were displayed. As shown in Figure 1, both the target and cursor appeared as solid vertical bars of equal height; however, the cursor bar was about twice as wide as the target bar. The cursor for the horizontal-axis tracking task was controlled by a set of pedals similar in location to the pedals used in an aircraft to control yaw. Pressing the left-hand pedal caused the cursor to move to the right; pressing the right-hand pedal caused the cursor to move left.

The target and cursor for the vertical-axis tracking task appeared in a rectangle located on the left portion of the display. The target and cursor used for the vertical-axis tracking task were the same size and shape as the target and cursor used for the horizontal-axis tracking task, but the bars were oriented horizontally (see Figure 1). The cursor was controlled with a single-axis joystick located to the left of the aviator in about the same position as a fixed-wing aircraft throttle. The single-axis joystick was identical to the dual-axis joystick and moved in a fore-aft and right-left directions; However, only fore-aft movement controlled cursor movement. The joystick's pivot point was located only a few inches below the aviator's hand. Movement of the joystick in the forward direction caused the cursor to move down; an aft movement of the joystick caused the cursor to move up. Aviators were instructed to use the single-axis joystick to keep the wide bar (cursor) aligned with the narrow bar (target).

To improve the statistical power of tracking performance data for comparisons of the different conditions used in the research, the characteristics of the dual-axis, horizontal, and vertical forcing functions were carefully controlled. For each tracking task the forcing function (the computer controlled movement of the cursor) was determined by the data stored in a parameter file. The contents of the three files have been combined and translated in Table 3. The files were generated in two stages. In the first stage, a pseudo-random series of integers from 3 to 10 were generated for each tracking task (dual-axis, horizontal, and vertical). Each series summed to the 180 and was used to define the number of seconds in discrete time periods during the 180 second trial in which the cursor would be forced in a particular direction. The first, third, and fifth columns of Table 3 contain the discrete time periods of the forcing function used in the experiment.

In the second stage, another pseudo-random series of -1s, 0s, and +1s was generated for each tracking dimension for each tracking task. These series of numbers were used to define the direction of cursor movement in the following manner. For horizontal tracking tasks -1 indicated a leftward movement, +1 indicated a rightward movement, and 0 indicated no horizontal movement. For vertical tracking tasks -1 indicated a downward movement, +1 indicated an upward movement, and 0 indicated no vertical movement. The second, fourth, and sixth columns in Table 3 contain the direction of cursor movement used for each tracking task.

The forcing function for each tracking task was exactly the same for each experimental trial. For each display frame generated, the computer program moved the cursor one pixel in the direction specified by Table 3. Thus, for the first three seconds of each trial in which dual-axis tracking was being performed, the cursor moved left one pixel and down one pixel 33 times a second; for the next 7 seconds, the cursor

Table 3

Forcing Functions for the Dual-Axis, Horizontal, and Vertical Tracking Tasks

Dual-Axis		Horizontal		Vertical	
Seconds	Direction	Seconds	Direction	Seconds	Direction
3	Left-Down	6	None	6	Up
7	None-None	6	Right	6	Up
7	Right-Down	4	Left	4	Up
5	None-Up	3	Left	3	None
4	Left-None	4	Left	5	Up
4	None-Up	3	Left	8	Up
5	Right-Down	4	None	5	Up
4	None-None	4	Left	3	None
6	Right-Up	5	Left	5	Down
3	Right-None	7	None	5	Down
4	None-Up	6	None	4	None
7	Left-None	7	Right	8	Up
6	None-Down	8	Left	5	None
4	Right-Down	6	None	8	Down
5	Right-None	3	Right	6	None
3	Left-None	5	Right	5	Down
8	None-Up	3	Right	7	None
6	Left-Down	8	Right	8	Up
6	Right-Down	5	None	7	None
8	None-Down	7	Right	6	Down
7	Right-Down	8	Right	6	None
8	Left-None	7	Right	6	Up
8	None-Down	4	None	3	None
5	Right-Down	4	Left	5	Up
6	None-None	5	Left	6	Up
4	Right-None	4	None	4	None
3	None-Up	6	Right	6	Down
8	Left-Down	6	None	7	None
5	Left-None	3	Left	8	Up
5	None-Up	5	None	8	Up
6	None-Up	7	Left	7	Down
10	Right-Up	7	None		
		10	Right		

was not forced in any direction. Although the aviators performed many repeated trials, the consistency of the forcing function was not apparent to the participants.

Task conditions. Seven conditions were constructed using the four basic tasks. Rather than simply assigning each task an arbitrary number, each condition is identified with a four character string; one character signifies the presence or absence of each of the four basic tasks. If the DL task was performed in the condition, the letter D is used in the first position. If the dual-axis tracking task was performed, the letter S (taken from the word stick) is used in the second position. Similarly, if the horizontal- and vertical-axis tracking tasks were performed, the letter P (taken from the word pedal) and the letter T (taken from the word throttle) are used in the third and fourth positions, respectively. The letter x replaces its characteristic letter when a task is not performed in the condition.

Using the coding scheme described above, the seven conditions created for this research are xSxx, Dxxx, DSxx, xSPx, DSPx, xSPT, and DSPT. The seven task combinations were selected from 24 possible combinations of the four tasks to provide a large variation in workload. Furthermore, if the tracking tasks are considered collectively as three levels of a single psychomotor variable, the selected conditions cover all possible combinations of a two level DL variable and a four level psychomotor variable (with the notable exclusion of the xxxx combination).

The DSPT condition requires the aviators to perform three tasks with two hands. The dual-axis tracking task, vertical tracking task, and the DL response entry required the use of the operator's hands. This condition was included with the expectation that the situation would lead to operator overload and therefore provide valuable information on pilot strategies and performance during overload conditions.

Model Development

To test the inter-analyst reliability and the robustness of the TAWL methodology, two analysts independently developed TAWL workload prediction models of the seven conditions. One analyst had 2 years experience in workload modeling and a PhD level education in experimental psychology; the other was a retired helicopter pilot who had 7 years experience in workload modeling and a masters level education. Each of the models were constructed with the TOSS software in accordance with the instructions given in Hamilton, Bierbaum, and Fulford (1991). The models are referred to hereafter as TAWL Model 1 and TAWL Model 2, respectively.

TAWL Predictions

At the end of the simulation of each condition, TOSS computed the peak, mean, and standard deviation for the half-second workload predictions. In addition, TOSS identified the intervals during which the performance of concurrent tasks resulted in excessive workload (referred to hereafter as overload). The three indexes of overload computed by TOSS and used in this research are described in the following paragraphs.

Component overload. A component overload occurs when the total workload for a single component reaches or exceeds a value of 8 during a half-second interval of the task simulation. Thus, several component overloads (i.e., cognitive, psychomotor, visual, etc.) could occur for each half-second interval. Because the maximum value on the 7-point workload component rating scales requires all the aviator's attention for that component, the value 8 was chosen as the overload threshold.

Overload condition. An overload condition exists during the period of time when at least one component overload occurs. A new overload condition is counted when the tasks contributing to a component overload change. Overload conditions identify the unique task environments that generate one or more component overloads.

Overload density. Overload density is the percentage of time during a condition that a component overload is present. Overload density is computed by dividing (a) the number of half-second intervals that contain component overloads by (b) the total number of half-second intervals in the condition.

Materials

Demographic questionnaire. The aviator demographic questionnaire (see Appendix A) was designed to collect personal, training, and flight experience data for the aviators who participated in the research. All aviators completed the questionnaire during the initial briefing.

NASA-Task Load Index (TLX). The NASA-Task Load Index (TLX) is a method that NASA developed to measure the subjective workload experience of an operator. NASA-TLX measures are generated by requiring the system operator to rate the mental demand (MD), physical demand (PD), temporal demand (TD), performance (PF), effort (EF), and frustration (FR) imposed by the system. The ratings, collected on

20-step bipolar scales, yield scores from 0 to 100 assigned to the nearest 5.

The subscale ratings are weighted for each individual and are combined into a single rating of workload. The individual weights are determined from a paired comparison task that is performed before the workload assessments are made. Paired comparisons require the operator to choose which subscale is more relevant to workload for each of the 15 unique pairs of the 6 subscales. The number of times a subscale is chosen as more relevant serves as the subscale weight for that operator. A TLX score from 0 to 100 is derived by multiplying the subscale rating for a task by the individual subscale weight, summing across subscales, and dividing by the total number of weights (15).

The three forms presented in Appendix B and a set of cards were developed from NASA Task Load Index (TLX): Paper-and-Pencil Version (1986). Pages B-2 and B-3 show the instructions that aviators were given on the use of the NASA-TLX rating scales. As suggested by the NASA pamphlet, the instructions were adapted for this research. Additionally, after consulting with the developer of the scales, the performance scale anchors were reversed such that good appeared at the right of the scale and poor appeared on the left; preliminary tests with the scales indicated that this scale arrangement corresponded more closely with raters' expectations.

Page B-4 presents the NASA instructions, entitled Sources of Workload. These instructions were used with a set of cards to collect the paired-comparison data needed to generate the individual scale weights. The 15 cards contained all possible pairs of scale titles.

In addition to the TLX, the equally weighted mean subscale workload (MSW) was computed. MSW is derived by computing the mean of the subscale ratings. The MSW is a less costly measure of subjective workload because the collection and use of the paired-comparison data is not required.

Aviators

Twenty AH-64A aviators (10 battle-rostered crews) from one battalion of an operational aviation regiment participated in this research. All crews completed all experimental sessions.

The demographic data for the 20 aviators indicate a range of experience that is typical of AH-64 operational units (see Table 4). The aviators had two distinctly different backgrounds: those with previous career experience in other helicopters (predominantly the AH-1) and those who proceeded directly from initial entry rotary wing training to the AH-64 Aviator Qualification Course (AQC). Across all indicators of experience, the pilots were more experienced than the copilot/gunners.

Apparatus

Computer test station. The computer test station used to present the conditions to the aviators consisted of an IBM compatible microcomputer with a 14-in color display that had a phosphor refresh rate of 60 Hz. The display was positioned approximately one meter from aviators' eyes. The computer was equipped with a Systems Research Laboratory Labpack board that performed the speech synthesis and analog-to-digital sensing of the tracking controls. The computer also was equipped with a separate numeric keypad that was used to collect aviator responses during dichotic listening trials.

Physiological recording equipment. Measures of aviator heart rate were collected using small silver/silver chloride

Table 4

Aviator Demographic Data at the Onset of the Research

Measure		Pilot (n = 10)	Gunner (n = 10)	Combined (n = 20)
Age (years)	median	30	24.5	29
	range	(24-35)	(22-30)	(22-35)
Months on active duty	median	93.5	81	76.5
	range	(55-187)	(25-130)	(25-187)
Months since AQC	median	39.5	17.5	26
	range	(12-56)	(0-28)	(0-56)
AH-64A flight hours	median	462.5	230	300
	range	(250-900)	(30-390)	(30-900)
Total flight hours	median	775	479.5	505
	range	(450-2300)	(190-600)	(190-2300)

Note. AQC = Aviator Qualification Course.

(Ag/AgCl) electrodes (3M Model 3320 Red Dot) and amplified using an EKG impedance pneumograph with an EKG amplifier (UFI Model RESP 1/EKG). To ensure good signals, the electrode contact impedance was measured using an electrode contact tester (UFI Model 1089 MK II Checktrode).

After the physiological heart signals were amplified, they were recorded along with synchronization signals generated by the computers using an 8 channel pulse-width-modulation data acquisition system (UFI Model 3370/8). The 8 channel system allowed the simultaneous recording of the heart signals from two aviators for later playback and analysis.

Procedures

In-briefing. Before beginning the first session, the experimenter briefed the participants on the purpose of the research and on the experimental procedures and tasks. After the briefing was completed, the aviators were instructed to complete the demographic questionnaires. Copies of the NASA-TLX instructions (see pages B-2 and B-3) were given to each aviator. The aviators were instructed to read the instructions carefully and to familiarize themselves with the meaning and use of the scales. Finally, the aviators were required to practice rating tasks using the scales.

Experiment procedures. Each aviator completed three identical experimental sessions over a six day period. The sessions were separated by at least one day. When the aviator arrived at the test facility, five physiological recording electrodes were applied to his skin. The two electrodes used to collect heart rate signals were placed along the mid-axillary lines (left and right) at about the level of the rib cage. A ground electrode was placed on the abdomen. The aviator was then seated at the computer test station, the electrode leads were connected to the electrodes, and the physiological recording was initiated.

Once the aviator had arranged the experimental apparatus in a comfortable configuration, the computer program controlling the experiment was initiated. The program presented each aviator with three trials of the seven tasks in the following order: xSxx, Dxxx, DSxx, xSPx, DSPx, xSPT, DSPT. Progress through the 21 trials was self paced. After each 3-minute trial, the aviators were presented with a brief description of the upcoming trial and were required to press a key on the numeric keypad to begin the next trial. Appendix C shows the computer file that controlled the program and contained the aviator instructions. After the third trial of each condition, the aviators were instructed to perform a NASA-TLX rating on the completed task.

Dichotic listening performance and tracking error were collected by the computer during all trials. Once the aviator had completed the 21 trials, he was disconnected from the electrode leads, the electrodes were removed, and he was thanked and dismissed for the day.

Data Analysis and Scoring

Tracking performance. The computer test station computed and stored tracking performance separately for each active tracking task. During each trial, the computer measured and stored the distance between the cursor and the target in centimeters 33 times a second. After each trial, the computer calculated the root-mean-squared (RMS) error for each tracking task.

Past research has shown that a subject's strategy can have dramatic effect on the performance of concurrent tasks (Gopher, 1980; Gopher, Brikner, & Navon, 1980; Schneider & Fisk, 1980; Wickens & Gopher, 1977). Some subjects may choose to devote about equal attention to each task being performed concurrently; others may devote far more attention to one task than to another, thereby trading poor performance on one task with good performance on another. An examination of the aviators' tracking performance suggested that tracking performance was, indeed, influenced by different coping strategies. Hence, for present purposes, it was not meaningful to examine performance separately on each of the three tracking tasks. Rather, the mean RMS error for concurrent tracking tasks was computed and used as a single index of tracking performance. For example, the index of tracking performance on condition xSPx was computed by summing the RMS for the two-dimension tracking task (S) and the horizontal tracking task (P) and dividing the sum by 2. In the case of the xSPT and the DSPT conditions, the index of tracking performance was computed by summing mean RMS error for all three tracking tasks and dividing by 3.

Dichotic Listening performance. During the DL task, aviators attended to the digits and letters presented to the designated ear and attempted to type the 9 digits that occurred in each DL stimulus string. Three measures of DL performance were calculated by the computer test station: errors of commission, omission, and exact match. Gopher and Kahneman (1971) and Gopher (1982) defined errors of commission and omission and used the measures to predict flying proficiency in high performance aircraft. Errors of omission are the number of digits present in the DL stimulus string that the aviator failed to enter. Errors of commission are the number of entries that were not present in the DL stimulus string. In both of Gopher's studies, errors of omission were the best predictors of success in flight training. The USAAVNC uses omission errors to aid in the assignment of aviators to particular aircraft after initial rotary wing training.

Another more demanding measure of DL performance was computed for each trial. An exact match error was counted anytime an aviator's input deviated from the DL stimulus string. The aviators that participated in this research are required to receive numeric strings over noisy radios during the normal conduct of their flight activities. The exact match measure was collected to eliminate the possibility of perfect performance (i.e., a floor effect) for errors of omission and commission. Additionally, errors of exact match can be computed using a much simpler scoring algorithm than is required to score the other types of errors.

Interbeat interval data correction. The heart rate signals were analyzed to determine the series of interbeat intervals of the heart using an IBM-compatible computer, a Data Translation Model 2808 analog-to-digital converter board, and the workload assessment monitor (WAM) software developed by the Human Engineering Group of the Armstrong Aerospace Medical Research Laboratory at Wright-Patterson Air Force Base, Ohio. Because of the high cost of analyzing heart rate variability data, only the data compiled on Day 3 were analyzed. The WAM software was used to analyze the heart rate recordings and to generate data files containing discrete lists of the interbeat intervals (IBIs) occurring during each trial.

Review of the output data files generated by the software indicated that it had missed some beats in most trials. Missed beats are indicated by a jump in the IBI interval to approximately twice the average. The failure of the analysis protocol to detect every heartbeat is not unusual in this type of research. Procedures for correcting missed heartbeats are necessary because including a single IBI with twice the average duration has substantial effects on estimates of the variability in the data.

For each trial that had missing data, the data were corrected using the following protocol. A running mean of the previous 30 IBIs was calculated. If an IBI was found that was 1.6 times the current mean, the IBI was divided by the integer that produced corrected IBIs closest to the IBI running mean. Consider the following example. The IBI running mean is 800. An IBI of 2500 is found in the data file. The 2500 IBI is divided by 2, 3, and 4 producing corrected IBIs of 1250, 833.33, and 625, respectively. The 2500 IBI is replaced with three IBIs of 833.33 because they are closest to the current running mean of 800. If WAM did not detect a beat for more than 5 seconds or if the number of missed beats exceeded 5% of the data, the trial was excluded from the analysis. The heart rate data for two aviators were eliminated from analysis because the number of trials rejected was too large. Of the remaining 18 aviators, only 9 of the 378 trials (18 aviators by 7 conditions by 3 trials) were rejected; in no case was more than one of the three trials for a condition rejected.

Time-based measures of heart rate variability. For each of the 18 aviators, 6 closely related time-based measures of heart rate variability were calculated and analyzed using methods described by Van Dellen, Aasman, Mulder, and Mulder (1985). The six measures are listed below:

- the standard deviation of the IBI,
- the coefficient of variation of the IBI,
- the root mean square of successive interval differences,
- the coefficient of variation of successive interval differences,
- the sum of absolute differences between successive IBIs, and
- the ratio of the IBI decelerations to the number of IBI fluctuations.

Although the heart rate variability literature indicates that any one of these measures would be sensitive to manipulation of task demands, each was computed and analyzed for comparison and evaluation.

Frequency-based measures of heart rate variability. Because there is some debate about the relative sensitivity of time-based versus frequency-based measures of heart rate variability (Van Dellen, Aasman, Mulder, & Mulder, 1985), a subset of 5 aviators' data was transferred to the Human Engineering Group of the Armstrong Aerospace Medical Research Laboratory at Wright-Patterson Air Force Base, Ohio, for further analysis. The IBI series were Fourier analyzed to measure the heart rate variability in two frequency bands. The low frequency band is correlated with the blood pressure control system and is centered around 0.10 Hz. The high frequency band is above 0.14 Hz. Changes in the high frequency component have been found to be correlated with changes in respiration and in the task repetition rate. Variability at the characteristic 0.10 Hz frequency for the blood pressure band corresponds to changes in the IBI series that rise and fall approximately every 10 s; variability at the characteristic 0.20 Hz frequency for the respiration band corresponds to changes in the IBI series that rise and fall every 5 s. These two frequency-based measures of heart rate variability were computed using the MXEdit software developed by Delta-Biometrics and the results were returned for subsequent analysis and plotting.

Statistical analyses. For each workload measure, a repeated-measures analysis of variance (ANOVA) was conducted to test its sensitivity to manipulations of two within-subjects factors: practice (day) and task demands (condition). The ANOVA assumption of homogeneity of population variance probably does not hold for distributions of TLX ratings, RMS tracking error, DL performance, or heart rate variability. Although violations of the homogeneity of variance assumption may not be critical with completely randomized designs, it can seriously affect the interpretation of F ratios in repeated-measures designs. In fact, the F test is known to be positively biased or about 2 or 3% more lenient than nominally set.

To solve the problem of heterogeneity of variance and to maintain the consistency of the tests across measures, the Geisser-Greenhouse correction for maximal heterogeneity (Giesser and Greenhouse, 1958) was used. The method assumes that the populations are not homogeneous, performs the usual analysis of variance, and evaluates the observed F ratios against new critical values. The method computes the F ratio in the classical manner and adjusts the p values assuming biased populations. When main effects for either factor were found, post-hoc one-way ANOVAs or contrasts were performed between adjacent levels of the factor to determine the level or levels where change occurred.

Results

The following four subsections describe the predictions generated by the two TAWL models and the results of the subjective, performance, and physiological workload measures. The two final subsections describe the relationships among these variables.

TAWL Workload Predictions

As previously described, two workload prediction models were developed using the TAWL methodology. The following three subsections describe the structure of each model and compare the models' structure and predictions.

TAWL Model 1. TAWL Model 1 has seven segments, seven functions, and seven tasks. Appendix D presents a complete representation of Model 1. Each condition was modeled using a segment decision rule. Four functions were defined to accomplish the tracking tasks. One function was used to control each of the three tracking tasks; the fourth function was used to visually monitor the screen. The remaining three functions were used to perform the DL task. One function was used to detect the beginning of the audio presentation; one function was used to monitor the auditory stimuli during presentation; and one function was used to enter the responses on the computer keypad. Seven tasks were used to construct the seven functions. One task was defined to accomplish the activities in each of the functions defined in the model.

Table 5 presents a summary of the workload predictions generated by Model 1. Two general workload metrics are shown in Table 5: the number of overload conditions (OC) and overload density (OD). Overload conditions were predicted for only two conditions: DSPx and DSPT. The overload density metric indicates that an overload condition is present during 40.8% of the DSPx condition and during 73.3% of the DSPT condition. Also shown in Table 5 are the average workload predictions for each of the workload components.

TAWL Model 2. TAWL Model 2 consists of seven segments, eight functions, and six tasks. Appendix E presents a complete representation of Model 2. Each condition was modeled using a segment decision rule. He defined four functions to accomplish the tracking tasks. One function was used to control each of the three tracking tasks; the fourth function was used to visually monitor the screen. The other four functions were used to accomplish the DL task; one function was used for each of the following:

- attend to the auditory stimuli during audio presentation,
- detect the beginning of the audio presentation,
- enter the responses to the first part of the DL string, and

Table 5

Workload Predictions for TAWL Model 1

Segment	OC	OD	AUD	KIN	VIS	COG	PSY
2: Stick Only	0	0.0	0.0	1.0	5.4	2.4	2.6
1: DL Only	0	0.0	3.9	0.4	1.5	4.6	0.9
3: DL and Stick	0	0.0	3.9	1.4	4.7	6.6	3.5
4: Stick and Pedal	0	0.0	0.0	2.0	5.4	3.6	5.2
5: DL, Stick, and Pedal	64	40.8	3.9	2.0	4.7	7.3	5.0
6: Stick, Pedal, and Throttle	0	0.0	0.0	3.0	5.4	4.8	7.8
7: DL, Stick, Pedal, and Throttle	126	73.3	3.9	3.0	4.7	8.5	7.6

Note. The order of the segments was changed to correspond with the order of task presentation in the experiment and with the order used in the other model. The following abbreviations are used: OC = Overload Condition, OD = Overload Density expressed as a percent of total segment time, AUD = Auditory, KIN = Kinesthetic, VIS = Visual, COG = Cognitive, PSY = Psychomotor, DL = Dichotic Listening.

- enter the responses to the second part of the DL string.

Six tasks were used to construct the eight functions. One task controlled the vertical and one task controlled the horizontal position of the cursors in the tracking functions. In addition, one task was defined to perform each of the following:

- monitor the computer screen,
- enter one DL response,
- monitor the DL stimuli during string presentation, and
- detect the beginning of DL string presentation.

A summary of the workload predictions generated by the Model 2 is shown in Table 6. Overload conditions were predicted only for the DSPT condition; an overload condition is present 28.9% of the time.

Comparison of TAWL models. Although the workload prediction models developed by the two analysts were similar, the models differed in two ways. First, there were differences in the structure of the two models. The top-down analysis of the conditions produced two different functional descriptions of the situation. These differences are not judged to be particularly important because, with one

Table 6

Workload Predictions for TAWL Model 2

Segment	OC	OD	AUD	KIN	VIS	COG	PSY
1: Stick (ST)	0	0.0	0.0	NR	5.4	1.0	2.6
2: Dichotic Listening (DL)	0	0.0	3.6	NR	1.4	3.4	0.6
3: ST and DL	0	0.0	3.6	NR	5.3	4.4	3.2
4: ST and Pedals (PD)	0	0.0	0.0	NR	5.4	2.0	5.2
5: ST, DL, and PD	0	0.0	3.6	NR	5.3	5.4	5.8
6: ST, PD, and Throttle (TR)	0	0.0	0.0	NR	5.4	3.0	7.8
7: ST, DL, PD, and TR	87	28.9	3.6	NR	5.3	6.4	8.4

Note. The following abbreviations are used: OC = Overload Condition, OD = Overload Density, AUD = Auditory, KIN = Kinesthetic, VIS = Visual, COG = Cognitive, PSY = Psychomotor, NR = Not Rated.

exception, both models produced the same sequence of tasks. The exception concerned the status of the tracking tasks during the numeric entry for the dichotic listening task. One analyst interrupted (stopped the performance of) the tracking tasks during numeric entry; the other analyst did not. The other differences between the models reflect subtle differences in the way the analysts performed the top-down analysis. Differences of this type illustrate that the TOSS software provides alternative methods for constructing functionally equivalent models. Most of the structural differences between the models did not contribute to differences in the models' predictions. However, the predictions were influenced by whether or not the model assumed concurrent performance of the tracking tasks and numeric entry tasks. In fact, the difference in assumptions about concurrent performance of tracking and numeric entry accounts entirely for the different number of psychomotor overloads predicted by the models.

The second way in which the models differed was in the analysts' workload ratings. For example, one analyst matched the cognitive component of the tracking tasks with the workload rating scale anchor "Alternative Selection" (1.2) and the other analyst matched it with "Simple Association" (1.0). Additionally, one analyst rated the cognitive component of the dichotic listening task with the scale anchor "Encoding/Decoding, Recall" (5.3) and the other analyst matched it with "Evaluation/Judgment (Consider Single Aspect)" (4.6). The differences in workload ratings account for the high number of cognitive overloads in Model 1 (segments 6 and 7), relative to Model 2.

The relationship between the two models was assessed in two steps. First, a Pearson product-moment correlation coefficient (*r*) was calculated for each workload component. The *r* values for the auditory, visual, cognitive, and psychomotor predictions were 1.00, .98, .99, and .99, respectively. Second, an overall workload prediction for each condition was derived by averaging the workload across components (see Figure 2). The *r* value for the overall workload predictions from the models was .99. The fact that the coefficients are so close to unity indicates that the predictions generated by the two models are highly related despite the differences in the models' structure.

Subjective Ratings

For each experimental session, aviators completed the NASA-TLX rating forms after finishing the third trial of each condition. The individually weighted sum of subscales (the TLX) and the equally weighted mean subscale workload (MSW) were analyzed. Analyses to compare the TLX and the MSW were conducted to evaluate the claim of Byers, Bittner, and Hill (1989) that the more easily computed MSW is nearly as effective as the TLX for assessing subjective workload.

Subscale analysis. Before the NASA-TLX could be used as a measure of TAWL validity, some analysis of the ratings' sensitivity to task demands was necessary. Additionally, the

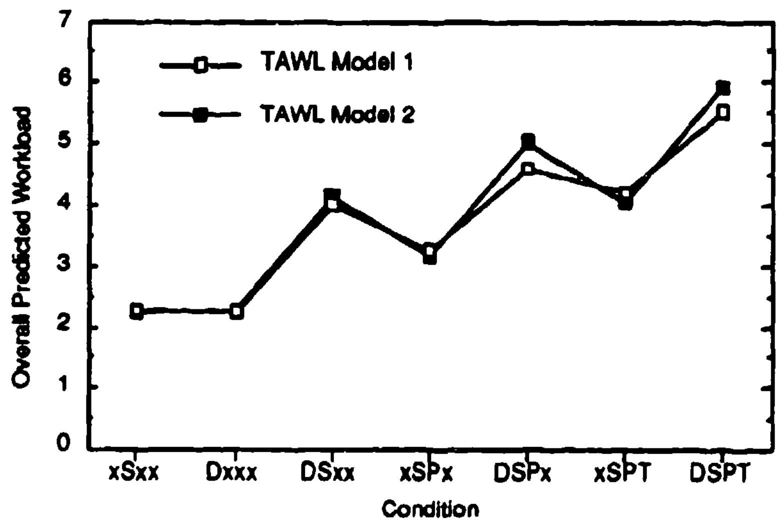

Figure 2. Overall predicted workload for Models 1 and 2 as a function of condition.

subscale results needed to be examined to confirm that the aviators used the scales appropriately. To begin these analyses, the NASA-TLX subscale ratings were pooled across aviators, means were calculated, and the means were plotted as a function of day and condition. As can be seen in Figure 3, ratings for most subscales varied systematically as a function of both day and condition. With only one exception (own performance), the ratings decreased as a function of day and increased as a function of condition. In contrast, the ratings of own performance did not vary as a function of day but decreased as a function of condition.

The error bars in Figure 3 depict the size of the standard error (plus and minus) of each mean. The error bars are presented in Figure 3 and in subsequent figures simply to provide an indication of the across-aviator variability in each data set.

An indication of the sensitivity and appropriate use of the subscales can be seen in the plot of physical demand (see Figure 3). The graph shows that the physical demand rating of the dichotic listening only (Dxxx) condition is considerably lower than the ratings for the other conditions. Presumably, this reflects the low physical demand of this condition relative to the other multi-task conditions and the single tracking task condition (xSxx). Without exception, the relationships between the NASA-TLX subscale values and both day and condition conform with expectations.

A repeated-measures ANOVA was conducted to test the sensitivity of each of the subscales to practice (day) and task demands (condition). Table 7 presents the results of the 18 F tests. Although some of the tests were significant and some were not, the Geisser-Greenhouse correction did not change the significance of any of the tests from the nominal F criteria. The effect of day was significant for all subscales except own performance; the effect of condition was significant for all subscales. A significant day by condition interaction was found for two subscales: mental demand and physical demand. The significant interactions indicate that the effect of practice differed as task demands changed.

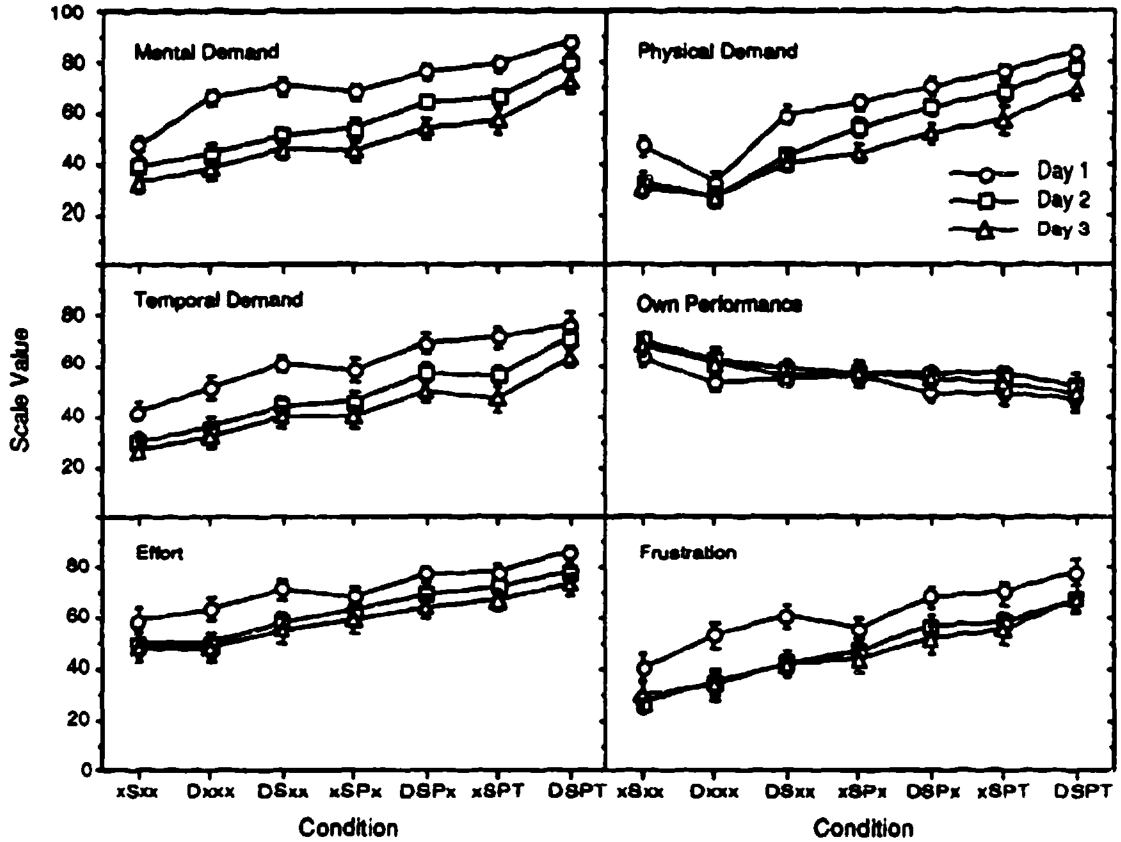

Figure 3. Ratings on six NASA-TLX subscales as a function of day and condition. (The error bars represent ± 1 standard error of the mean.)

<u>Task load index and mean subscale workload analysis</u>. With evidence that the aviators used the subscales appropriately and that the subscales were sensitive to the experimental manipulations, the NASA-TLX and the MSW were computed and plotted (see Figure 4). As can be seen in Figure 4, the ratings varied systematically as a function of both day and condition; the ratings decreased as a function of day and increased as a function of condition.

Repeated-measures ANOVA tests were conducted to test TLX and MSW sensitivity to practice (day) and task demands (condition). The last two rows Table 7 present the results of the 6 F tests. Both measures showed significant differences as a function of day and condition and the TLX measure showed a significant interaction between day and condition. However, when the Geisser-Greenhouse correction

Table 7

F Ratios Resulting From ANOVA Tests of NASA-TLX Sensitivity to Day and Condition

	Variable		
Scale	Day $df = 2,38$	Conditions $df = 6,114$	Day by condition $df = 12,228$
MD	26.79**	78.64**	3.76**
PD	13.46**	120.39**	3.03**
TD	16.15**	52.32**	1.17
OP	2.50	7.01**	0.75
EF	15.05**	24.61**	1.00
FR	11.25**	73.28**	1.45
TLX	23.73**	95.57**	2.84*
MSW	23.03**	104.19**	2.22

Note. The following abbreviations are used as column headings: MD = Mental Demand, TD = Temporal Demand, OP = Own Performance, EF = Effort, FR = Frustration, TLX = Task Load Index, and MSW = Mean Subscale Workload.

*Adjusted values of $p < .05$. **Adjusted values of $p < .01$.

was used, the interaction was not significant for the MSW. In summary, the results indicate that both measures decrease significantly as a function of practice and increase significantly as a function of task demand.

Scale relationships. Pearson product-moment correlation coefficients (r) were computed to assess the relationships among the subscales, the TLX, and the MSW. Only the data for the final day of practice were analyzed for this purpose. Table 8 shows the results of the analyses. The correlations among five of the six subscales were found to be high ($.42 \leq r \leq .83$) and positive. In contrast, the relationship between own performance and all other subscales was found to be either negative and low ($.13 \leq r \leq .30$) or not statistically significant. These findings are logically consistent.

Specifically, the findings show that (a) conditions that are demanding with respect to one dimension (e.g., mental demand) are also demanding with respect to other dimensions (e.g., physical demand and temporal demand), (b) conditions that are the most demanding tend to require the most effort and create the most frustration, (c) assessments of own performance tend to decrease as task demands increase, and

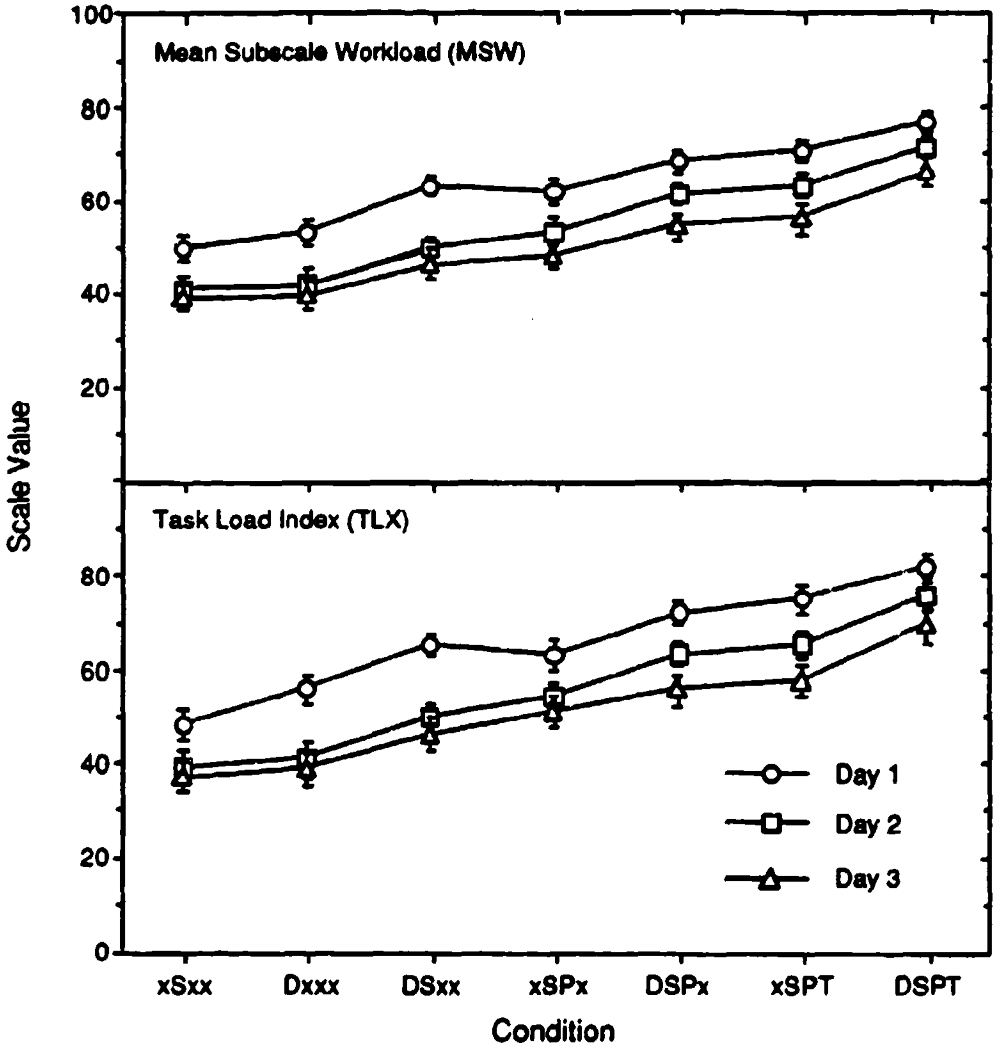

Figure 4. The Task Load Index and mean subscale workload as a function of day and condition. (The error bars represent ± 1 standard error of the mean.)

(d) as task demand increases, greater effort does not necessarily result in better performance.

The correlation between the individually weighted TLX and the equally weighted MSW was found to be positive and very high (r = .96). The powerful relationship between the measures strongly supports Byers' claim that MSW is nearly as effective as TLX for assessing subjective workload (Byers et al., 1989). The similarity of the two measures also is indicated by the similarity in the strength of their

Table 8

Correlations Among the NASA-TLX Subscale, the TLX, and the MSW Ratings

Scale	MD	PD	TD	OP	EF	FR	TLX
PD	.83*						
TD	.80*	.71*					
OP	-.23*	-.21*	-.13*				
EF	.65*	.61*	.55*	-.00			
FR	.69*	.64*	.69*	-.30*	.42*		
TLX	.91*	.85*	.84*	-.19*	.71*	.83*	
MSW	.91*	.87*	.87*	-.02	.76*	.78*	.96*

Note. For all correlations $N = 420$. The following abbreviations are used as column headings: MD = Mental Demand, TD = Temporal Demand, OP = Own Performance, EF = Effort, FR = Frustration, TLX = Task Load Index, TLX = Task Load Index, and MSW = Mean Subscale Workload.
*$p < .01$.

relationships with the individual subscales. Table 8 shows that, for five subscales, the subscales' correlation with TLX is nearly identical to the corresponding subscales' correlation with MSW; the correlation coefficients for corresponding subscales differ by ±.05 or less. The correlation between own performance and TLX is negative and statistically significant ($r = -.19$); the correlation between own performance and MSW also is negative but is not large enough to reach statistical significance ($r = -.02$).

The relationships quantified in Figures 3 and 4 and in Tables 6 and 7 provide convincing evidence that the aviators who participated in this study used the TLX rating scales as they were designed to be used, and that the resulting ratings provide a sensitive measure of workload that can be used as a criterion in validating the TAWL methodology.

Performance Measures

The computer test station produced objective measures of tracking and dichotic listening performance. These two measures were analyzed to confirm that they were, in fact, sufficiently reliable and sensitive to be used as criteria for validating the TAWL methodology. The results of the performance measures analyses are presented in the following two subsections.

Tracking performance. Figure 5 shows RMS error as a function of day and condition. The curves in Figure 5 support the assumption that the task conditions employed in this study present different demands on the aviators. It can be seen in Figure 5 that (a) RMS is higher for the two-task condition (DSxx) than for the single-task condition (xSxx), (b) both three-task conditions have higher RMS error than either two-task condition, and (c) the four-task condition has higher RMS error than either three-task condition.

The size of the increment in RMS error that results from adding a single task depends on the task that is added. Adding a second tracking task (xSxx vs. xSPx) results in a larger increment in RMS error than adding the dichotic listening task (xSxx vs. DSxx). However, adding a third tracking task results in about the same increment in RMS error as adding a DL task to the same two-task condition. The largest increment in RMS error results from adding a fourth task to a three-task condition; the addition of a third tracking task to Condition DSPx results in about the same increment in error as adding the dichotic listening task to Condition xSPT. Despite the attempt to use the four-task condition (DSPT) to study operator strategies under overload

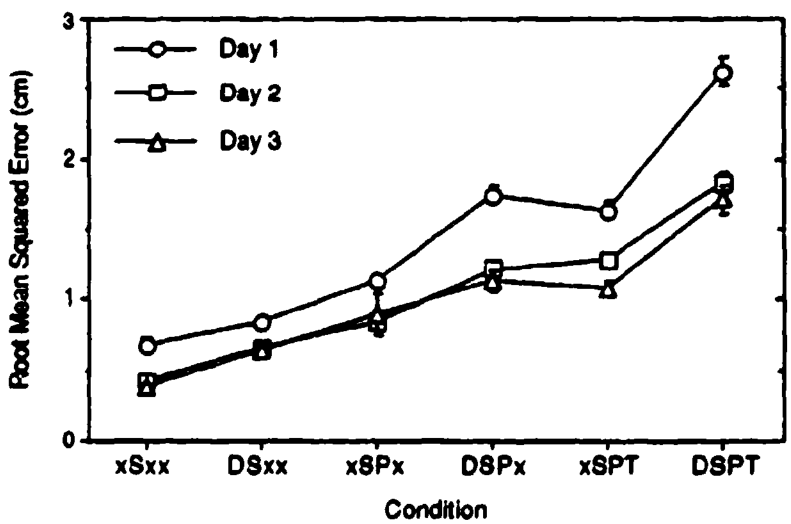

Figure 5. Tracking performance as a function of day and condition. (The error bars represent ± 1 standard error of the mean.)

conditions, performance on that condition did not degrade catastrophically as was expected.

A comparison of the RMS error curves for the three days indicates that although practice on Day 1 resulted in a substantial decrease in error on Day 2, no further improvement in performance resulted from the practice on Day 2. It is important to note, however, that the relationships between RMS error and condition described above did not change with practice on either Day 1 or Day 2.

Repeated-measures ANOVA tests were conducted to test the effect of practice (day) and task demands (condition) on tracking performance. Practice effected tracking performance significantly, F (2, 38) = 41.03, p < .0001; the difference between Day 1 and Day 2 was significant but the difference between Day 2 and Day 3 was not. Task demands also effected tracking performance significantly, F (5, 95) = 103.09, p < .0001. With only one exception, conditions were different from the preceding ones; condition DSPx did not differ significantly from condition xSPT.

The aviators employed different strategies to cope with the task demands (both physical and mental) of the four-task condition (DSPT). All 20 twenty aviators employed some strategy to maintain performance; none were observed to ignore or shed one of the tasks. The two most common strategies were (a) to control the vertical tracking joystick with the left forearm while entering DL responses with the left hand, and (b) to control the vertical tracking joystick with the left hand, the dual-axis joystick with the right hand, and enter the DL responses with the right hand intermittently.

<u>DL performance</u>. The dichotic listening task was included in four of the seven conditions used in the research. Figure 6 presents the number of DL errors of commission, omission, and exact match as a function of day and condition; in all cases, errors were averaged across aviators. As can be seen in Figure 6, a floor effect was not evident in any of the data; all three measures showed higher error rates as task demand increased. The errors of commission were less orderly and more variable than either omission or exact match errors. The trends for omission and exact match errors appear quite similar.

Other studies of dichotic listening have found dichotic listening performance to be better with the left than the right ear (Bryden, 1969; Gopher & Kahneman, 1971; Kimura, 1967; Treisman & Geffen, 1968). Error rates were higher when

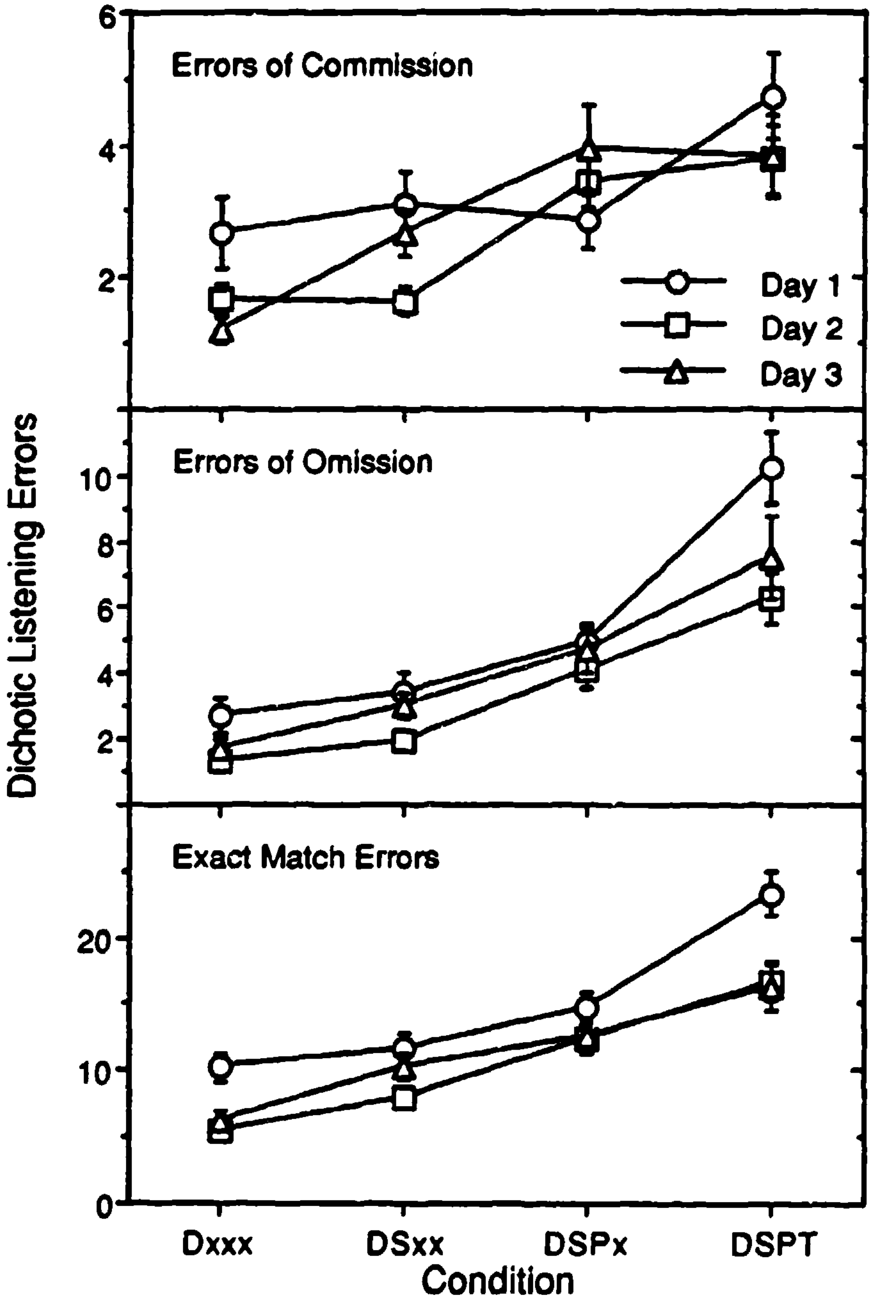

Figure 6. Dichotic listening performance as a function of day and condition. (The error bars represent ± 1 standard error of the mean.)

the relevant stimuli were presented to the left ear. The mean per-stimulus error rates for errors of commission, omission, and exact match for the left and right ear were .40 (L) vs. .34 (R), .62 (L) vs. .45 (R), and 1.67 (L) vs. 1.39 (R), respectively.

Three repeated-measures ANOVA tests were conducted to test the effect of practice (day) and task demand (condition) on DL performance. Table 9 shows the size and statistical significance of the 9 *F* ratios generated by the ANOVA tests. The effect of day was significant for all measures except errors of commission. The effect of condition was significant for all measures. None of the measures showed a significant day by condition interaction. For both omission and exact match error, the difference between Day 1 and Day 2 was significant but the difference between Day 2 and Day 3 was not.

In summary, both objective measures used in this research indicate that (a) practice after Day 1 did not result in improved aviator performance, and (b) error increases systematically as a function of task demand throughout the range of conditions used.

Physiological Measures

Eight time- and frequency-based measures of heart rate variability were collected for each experimental trial. The measures' sensitivity to manipulations of task demand were analyzed to determine their utility as criteria for validating the TAWL methodology. The results of the analyses are presented in the following three subsections.

Table 9

F Ratios Resulting From ANOVAs Performed to Test the Effect of Day and Condition on Dichotic Listening Performance

	Variable		
Error Measure	Day *df* = 2,38	Conditions *df* = 3,57	Day by condition *df* = 6,114
Commission	1.59	7.82**	2.01
Omission	4.62*	19.47**	1.49
Exact Match	8.62**	27.99**	1.60

*Adjusted values of $p < .05$. **Adjusted values of $p < .01$.

Time-based measures. The analysis of heart rate and heart rate variability was begun by computing each aviator's average IBI for each condition. Then the mean IBIs were pooled across aviators and a grand mean IBI was calculated for each condition; the grand mean IBI for each of the seven conditions is shown in Figure 7. As can be seen in Figure 7, the IBIs were longest for the xSxx condition; shorter for the Dxxx and DSxx conditions; and shortest for the xSPx, DSPx xSPT, and DSPT conditions.

A repeated-measures ANOVA showed a significant effect for condition, F (6, 102) = 10.60, $p < .0001$; however, the only significant difference between successive points was between the DSxx condition and xSPx condition. IBI, in itself, was not expected to be sensitive to manipulations of task demands; the measure is presented here to aid in the interpretation of heart rate variability described in the following paragraphs.

The six time-based measures of heart rate variability described in the methods section were computed, analyzed, and plotted as a function of condition. For ease of comparison, the data for all six measures are shown in Figure 8.

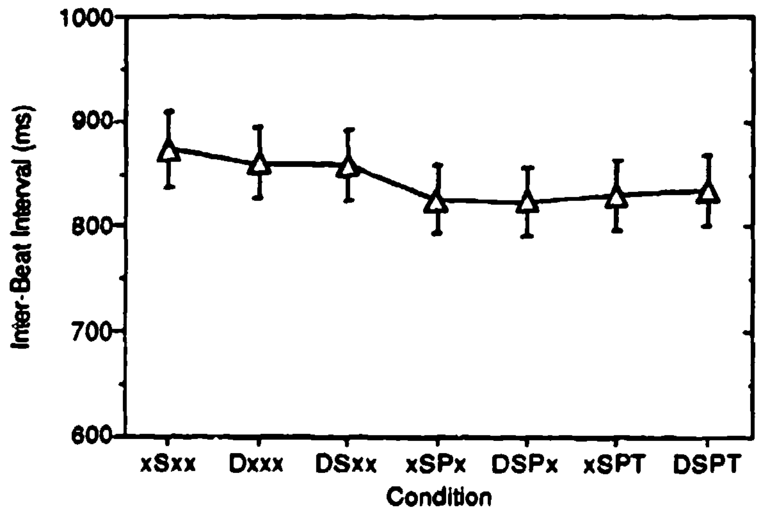

Figure 7. Mean heart interbeat interval as a function of condition. (The error bars represent ± 1 standard error of the mean.)

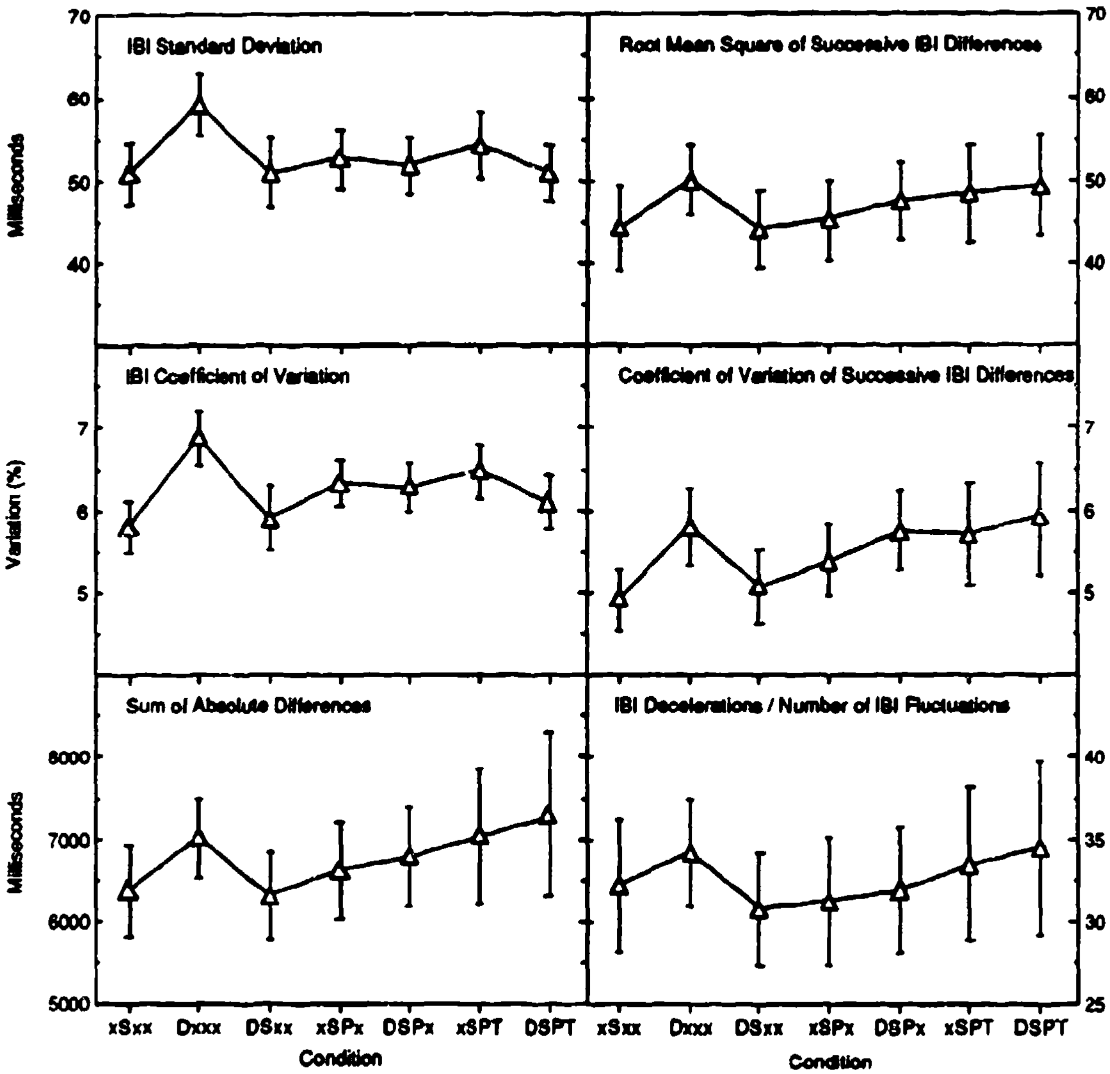

Figure 8. Six measures of heart rate variability as a function of condition. (The error bars represent ± 1 standard error of the mean.)

The data in Figure 8 must be interpreted in light of the fact that measures of heart rate variability are expected to decrease as task demand increases (Aasman et al., 1987). An examination of the curves in Figure 8 shows that none of the 6 time-based measures of heart rate variability conform with expectations. To the contrary, the plots of four measures suggest that heart rate variability increases as task demand increases.

Figure 8 shows that condition Dxxx produced a very high degree of heart rate variability relative to most other conditions. For two measures (IBI Standard Deviation and IBI Coefficient of Variation), heart rate variability was higher for condition Dxxx than for any other condition; for the remaining four measures, heart rate variability was higher for condition Dxxx than for most other conditions.

A repeated-measures ANOVA was conducted for each of the six time-based measures of heart rate variability. A significant condition effect was found only for IBI standard deviation, $\underline{F}$ (6, 102) = 2.36, p < .05, and IBI coefficient of variation, $\underline{F}$ (6, 102) = 2.55, p < .05. For both measures, the Dxxx condition differed from the xSxx condition and DSxx condition but no other successive contrast was significant. Hence, although two measures of heart rate variability vary significantly as a function of condition, the relationship is opposite to the findings in the literature. The cause of the unexpected relationships is unknown.

Frequency-based measures. The first step of the frequency analysis of heart rate variability eliminates approximately 20 IBIs from the beginning and end of each trial. Mean IBI was calculated for the remaining data and is presented in Figure 9 to aid in the interpretation of the measures of heart rate variability described in the following paragraphs.

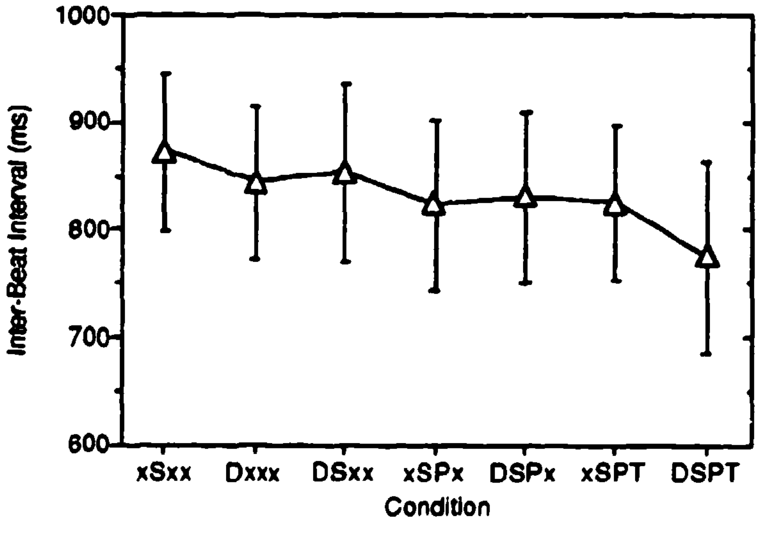

Figure 9. Heart interbeat interval as a function of condition. (The error bars represent ± 1 standard error of the mean.)

The frequency-based analysis measured the variability of the IBI series in two frequency bands. The low frequency or blood pressure band ranged from 0.06 and 0.14 Hz. The high frequency band or respiration band ranged from 0.15 to 0.40 Hz. Figure 10 presents the results of these analyses. In general, the trends for both measures were similar. The respiration band had a greater amount of variance than the blood pressure band. Both measures showed high heart rate variability for the Dxxx condition. However, the effects of condition were not significant for either of the measures.

<u>Summary</u>. Most time- and frequency-based measures of heart rate variability were not sensitive to manipulations in task demands. When the measures showed significant differences between conditions, the trends in the data were clearly different from the trends apparent in the performance and subjective measures. For these reasons, measures of heart rate variability were judged to be of little value as criteria for validating the TAWL methodology. Although, heart rate variability is not discussed further in the main body of this report, interested readers can find the heart rate variability statistics in the Appendix F.

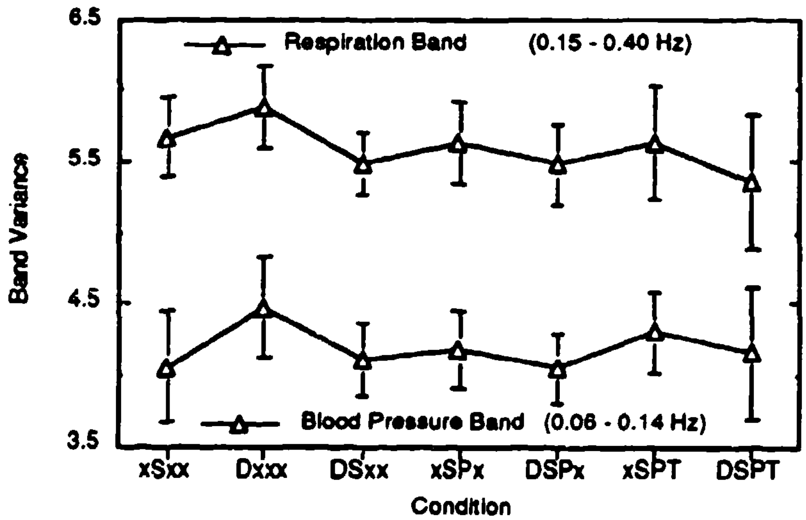

Figure 10. Two frequency-based measures of heart rate variability as a function of condition. (The error bars represent ± 1 standard error of the mean.)

Relationships Among Measures

The relationships between individual measures and conditions (task demands), described in previous sections, are the relationships that have the greatest bearing on the selection of the measures that are to serve as criteria for assessing the validity of the TAWL methodology. However, it also is worthwhile to examine the relationship among the measures themselves. It is of particular interest to determine the intercorrelation among the measures that were found individually to be highly correlated with task demand. A low intercorrelation among measures (that are each highly correlated with task demand) would suggest that task demand is multidimensional and that the measures serve to quantify two or more of the different dimensions. Conversely, high intercorrelations among measures indicates that task demand, as defined in this study, is unidimensional and can be quantified effectively with several different measures.

Appendix F is a comprehensive intercorrelation matrix that shows the Pearson product-moment correlation coefficient (r) between every possible pair of measures investigated in this study. Although many of the correlation coefficients shown in Appendix F have no direct relevance for this study, a comprehensive intercorrelation matrix was developed with the expectation that the data may be of value to other researchers who have an interest in workload measurement. All correlation coefficients shown in Appendix F and all correlation coefficients discussed below were computed using only the data compiled during the last practice day because aviators' proficiency was highest and their performance most stable on the final day.

Table 10 shows the intercorrelation among five measures that, individually, were highly correlated with task demand. It can be seen in Table 10 that errors of commission is the only measure that was not found to be highly correlated ($r \geq .985$) with every other measure. Although the correlation coefficients between errors of commission and the other four measures varied from .789 to .912, they were not large enough to reach statistical significance at the .05 level because of the small number of pairs in the data set (n = 3 or 4). The high correlation among RMS tracking error, DL omission errors, DL exact match errors, and TLX leave little doubt that these measures are assessing the same thing. It is possible that dichotic listening errors of commission measure a different dimension than the other four measures. However, a more likely explanation is that errors of commission are a less effective measure than the other four.

Table 10

Correlation Coefficients Among Effective Workload Measures

Measure		RMS	OMISS	COMISS	EXERR	TLX
OMISS	*r*	.996				
	p	.053				
	n	3				
COMISS	*r*	.789	.846			
	p	.422	.154			
	n	3	4			
EXERR	*r*	.996	.981	.912		
	p	.059	.019	.088		
	n	3	4	4		
TLX	*r*	.990	.999	.867	.988	
	p	.000	.001	.133	.012	
	n	6	4	4	4	
MSW	*r*	.985	.999	.871	.988	.995
	p	.000	.001	.129	.012	.000
	n	6	4	4	4	7

Note. The following abbreviations are used as headings: RMS = Root Mean Squared Tracking Error, OMISS = Dichotic Listening (DL) Errors of Omission, COMISS = DL Errors of Commission, EXERR = DL Exact Match Errors, TLX = NASA Task Load Index, and MSW = Mean Subscale Workload.

Relationship Between TAWL Predictions and Criterion Measures

The results presented previously show that the TAWL models generated workload predictions that varied systematically and in an expected manner with the seven conditions presumed to represent different levels of task demand. Furthermore, results were presented showing that both subjective workload measures and performance measures (but not physiological measures) varied systematically and in an expected manner with conditions. Given these results, it is meaningful to assess the validity of the TAWL methodology by determining the extent to which the TAWL predictions are related to the subjective and performance measures that have been shown to be highly correlated with conditions. Specifically, it is assumed that statistically significant correlations between the TAWL predictions and the measures used as validation criteria supports the conclusions that the TAWL methodology generates valid predictions of operator workload.

The relationship among TAWL predictions and selected criterion measures are shown in Table 11. Correlation coefficients are shown for all but one of the combinations of (a) 6 TAWL component predictions (visual, auditory, kinesthetic, cognitive, and psychomotor) and the average of the component predictions, and (b) two performance measures (RMS tracking error, DL exact match error) and one subjective rating measure (TLX workload index). Because the TAWL predictions for the auditory component were constant for all DL conditions, it was not possible to compute a correlation coefficient between the TAWL auditory component predictions and DL exact match errors. Errors of commission was excluded as a criterion measure because it was judged to be less reliable than the other two measures of dichotic listening. Errors of omission and TLX mean subscale workload were excluded only because of their very high correlation with exact match errors and NASA TLX, respectively.

Examination of Table 11 shows that the size and statistical significance of the correlation coefficients

Table 11

Correlation Coefficients Among TAWL Predictions and Criterion Measures

			TAWL prediction				
Measure	*n*	Model	VIS	AUD	KIN[a]	COG	PSY
RMS	6	1	-.450	.450	.872*	.760	.856*
		2	-.450	.450	NR	.778	.919*
EXERR	4	1	.802	NC	.999*	.992*	.999*
		2	.802	NC	NR	.996*	.995*
TLX	7	1	.367	.189	.921*	.738	.909*
		2	.458	.189	NR	.708	.938*

Note. The following are used as headings: VIS = Visual , AUD = Auditory , KIN = Kinesthetic, COG = Cognitive , PSY = Psychomotor , AVE = Average, RMS = Root Mean Squared Tracking Error, EXERR = Dichotic Listening (DL) Exact Match Errors, TLX = Task Load Index, and NR = Not Rated. The Auditory predictions were constant for DL trials; Correlations could not be computed (NC).

[a]Kinesthetic ratings were not assigned in TAWL Model 2.

*$p<.05$

tend to be consistent across the two models and the three criterion measures. However, the size of the correlation coefficients vary substantially across TAWL component predictions.

The TAWL visual component predictions are not significantly related to any of the three measures. This result can be attributed to the fact that the model only predicted three levels of visual workload for all the conditions. The lowest level was predicted for the Dxxx condition. The highest level was predicted for the three tracking only conditions (xSxx, xSPx, xSPT). The intermediate level was predicted for the mixed (DL and tracking) conditions that required the aviators to switch visual attention from the relatively low visual workload task of finding the numbers on the keypad to the high workload task of continuous target tracking. The switching of visual attention from the tracking task to the numeric entry task reduces visual workload and results in greater RMS tracking error. Thus, the negative relationship between the TAWL visual component predictions and RMS tracking error is logically consistent.

The TAWL auditory component predictions were not found to be significantly correlated with either of the two measures for which correlation coefficients could be computed. As was suggested for the visual component predictions, the lack of a statistically significant relationship between the measures the TAWL auditory component predictions is due to the lack of variability in the demands of the auditory task across conditions. That is, although overall workload varied across conditions, the model predicted one level of auditory workload when the condition required the dichotic listening task and a second level when the condition did not require the performance of the dichotic listening task.

As has been found in previous research, when kinesthetic and psychomotor ratings are used in TAWL models, the workload predictions generated for each component are so closely related that they can be considered redundant. The similarity of the two component predictions is indicated by the similarity in the strength of their relationships with the three measures. Table 11 shows that the three measures' correlation with the TAWL psychomotor component predictions is nearly identical to the corresponding measures' correlation with the TAWL kinesthetic component predictions; the correlation coefficients for corresponding measures differ by ±.02 or less. Nonetheless, of all the TAWL workload components, the psychomotor and kinesthetic predictions show the strongest relationship with the criterion measures.

The cognitive predictions were most closely related to the DL measure and were not found to be correlated significantly with either RMS errors or TLX workload ratings. In both cases, however, the correlation coefficients were nearly large enough to reach statistical significance.

The average TAWL predictions, similar to the psychomotor component predictions, show a consistently strong relationship with each of the criterion measures. The finding that predictions for the average TAWL predictions are related to the criterion measures is logically consistent with the manner in which task demand varied across conditions.

In summary, the TAWL workload predictions from both models show strong relationships with the subjective, and performance workload measures used as criteria for TAWL validity. Although some differences between models produced by different analysts were apparent, these findings strongly support the validity of the TAWL workload prediction methodology.

Discussion

The high correlations between the TAWL model predictions and the measures of workload support the conclusion that the TAWL methodology generates valid predictions of operator workload. However, several issues raised by the results of the research are addressed before stating the final conclusions. The discussion is organized into five subsections. The first subsection discusses the differences between the TAWL models produced by different analysts. The second subsection reviews the concept of overload as defined in the TAWL methodology in light of the aviator performance. The third subsection reviews the limitations of the current research design and the fourth subsection makes recommendations for further research. The final subsection summarizes the major conclusions of the project.

TAWL Models Constructed by Different Analysts

Differences in the models generated by different analysts for the same task conditions warrant careful review. If a methodology does not constrain the analyst sufficiently to produce valid models, the method itself can never be validated, only the individual models. However, the requirement that a methodology constrain analysts sufficiently to produce exactly the same model is probably unreasonable. Although some inter-analyst differences are to be expected, the differences cannot reduce or change the nature of the relationship between the model's predictions and the criterion measures. Described below are three types of differences between the models that were identified.

Structural differences. As described above, the TAWL methodology uses a top-down analysis to define the logical structure of the actions of system operators. The output of the analysis is a list of the tasks that operators must perform. In the process of identifying the individual tasks in a system, at least two other levels of analysis are described (e.g., segments, functions).

Defining these intermediate levels aids analysts in at least two ways. First, the levels link the individual tasks with the global goals of the system. Second, they help the analyst manage the complexity of the analysis. Analyzing system use at different levels greatly reduces demands on an analyst's working memory while allowing for the management of a large amount of complexity. For example, it is a simple matter for an analyst to describe the sequencing and interactions of the 5 to 15 functions in a segment; it is far more difficult to describe the sequencing and interactions of the 50 to 100 tasks in the same segment.

Some of the differences between the models generated for this research were differences in the structure of the intermediate levels. Because only the task level of the analysis is directly observable as operator actions, it is difficult to verify the validity of higher levels of the analysis. In other words, analysts may differ in the manner in which they group tasks into functions; however, if two different functional groupings of tasks produce the same sequence of task performance, there is no criterion with which to determine the veracity of the possible candidates. Although operators probably use logical groupings of tasks that might be determined empirically, the exact structure of these intermediate levels may not affect the predictions generated by the model.

The analyst is only required to define the intermediate levels in a way that generates valid task sequencing for the operation of the system. Only the tasks determine TAWL workload predictions; analysts do not assign ratings directly to either the functions or segments. Thus, if one analyst considered a set of 16 tasks to be one function and another analyst considered the tasks to be two functions, the validity of models would not be compromised if each analyst used the functions to generate equivalent task sequencing.

This discussion highlights the importance of one of the procedures recommended by the TAWL methodology. The methodology described by Hamilton, Bierbaum, and Fulford (1991) recommends that subject matter experts (SMEs) review and approve the final task sequencing generated by the simulation of segments. In conclusion, differences in the structure of the intermediate levels of TAWL models may be observed without necessarily compromising the validity of the models so long as they produce valid task sequences.

Rating differences. The second type of difference between the models were rating differences. After defining the task sequences, analysts use workload rating scales to rate the attentional demands of each task. The ratings are made for components that describe the sensory, cognitive, and psychomotor demands of tasks. The predictions generated by the methodology are the within-component sums of these ratings. Thus, differences in the assigned ratings between analysts directly change the predictions generated by the models.

Rating differences between analysts appear to be the most serious threat to the validity of the method. However, the differences in the predictions generated by the models developed for this research did not substantially reduce the validity of either model. Apparently, the consistent use of the rating scales by both analysts was sufficient to maintain the validity of the predictions. This is surprising because the workload predictions generated by models of low complexity are more highly dependent on each rating than those generated by models of high complexity.

A relatively small number of ratings formed the basis for the predictions generated by the models used here. The workload predictions of Model 1 and Model 2 were dependent on a total of 35 and 24 ratings, respectively. Examples of the number of ratings in typical TAWL applications range from 828 in the UH-60A model (Bierbaum, Szabo, & Aldrich, 1989) to 3540 in the AH-64A model (Hamilton & Bierbaum, 1992). Large numbers of ratings reduce the individual impact of a single rating.

The fact that rating differences did not seriously affect the validity of the models investigated does not argue for the haphazard use of the rating scales. Indeed, consistent use of the scales is assumed to have maintained the validity of the models. Whenever possible, conclusions should not be drawn from comparisons of predictions generated from models produced by different analysts.

Finally, TAWL methodology recommends that the workload ratings be made by the consensus of more than one analyst. This procedure improves the consistency and accuracy of the ratings within a single model. Furthermore, consensual rating may allow for the comparison of TAWL predictions generated by different groups of raters because group rating would be expected to reduce individual rating bias.

Opinion differences. Differences of opinion were the last type of difference between the models. The analysts constructed different models to reflect their opinions about how the aviators performed the tasks. One analyst believed that the aviators interrupted the tracking tasks while they entered the DL responses on the keypad. Because DL response entry occupied such a large proportion of trial time, the other analyst assumed that task sharing would have to occur and he did not interrupt the tracking tasks during DL response entry.

The fact that this difference in opinion did not lead to differential validity of the models is surprising. The total number of opinion based decisions made in developing the models for this research was small; each had the potential to change substantially the predictions generated by the model. Again, the best approach to minimizing the potential effects of differences in analyst opinions lies in the consensual development of TAWL models.

The Concept of Overload in the TAWL Methodology

The concept of operator overload has played a key role in the TAWL methodology since its initial development. The notion that operators can maintain good performance on tasks until they reach a point of catastrophic failure was popular in information processing and workload literature when TAWL was originally developed. The concept is intuitive, is supported by laboratory research, and is useful in illustrating the detrimental effects of workload on performance.

More recently, however, the workload literature has begun to recognize that good performance followed by catastrophic failure is only one of many different ways that operators may respond when presented with high task demands. Examination of the entire RMS tracking and DL performance functions here did not reveal evidence of a change in performance that would indicate the existence or level of an overload threshold, even though both models predicted overloads in the multi-task conditions. Clearly, the current findings do not support the catastrophic failure concept of workload.

The concept of operator overload was useful initially in the investigations of single pilot operation of scout/attack helicopters (Aldrich et al., 1984; McCracken & Aldrich, 1984) because of the unusually high demands of that situation. However, in subsequent investigations of production helicopters, overload conditions were rarely observed and other, more sensitive, measures of workload (e.g., average component workload) were used to accomplish the objectives of the research.

Synopsis. Although the concept of task overload is useful when thinking about performance on individual tasks, the concept of a fixed level of task demands at which performance degrades has limited utility for several reasons. First, its genesis was in the analyses of a very high workload situation where the researchers evaluated different design options. There are other measures of task demands (e.g., mean workload) that show better measurement qualities (e.g., sensitivity) for a broader range of task environments and equipment manipulations. Second, although the method used the overload concept, it was arbitrarily placed at 8 and it has never been empirically determined. Finally, there is growing evidence that the relationship between task performance and workload is complicated and cannot be explained using any single concept. As such, future applications of the TAWL methodology should reduce reliance on the concept of operator overload and continue to measure and report on all available information describing the operator's task loading and performance strategies.

Limitations of the Research Design

There are at least two aspects of the current research design that limit the ability to generalize the results. The first and most fundamental problem was the artificial task environment used in the research. To the merit of the design, the tasks were selected to have as much relevance to the operational rotorcraft environment as possible and to present the aviators with a broad range of workload. Unfortunately, the simplicity of the environment was apparent in the lack of complexity of the resulting TAWL models, especially with comparison to the environments in which TAWL has typically been applied. Until a TAWL model with more typical complexity is validated, the validity of complex TAWL models cannot be assumed.

Another aspect of the design may confound the findings of the research. Because access to the aviators was limited and because the aviators were unfamiliar with the experimental tasks, the experimental sessions were designed to maximize the level of proficiency obtained in the time allotted. Thus, the sequence of conditions was fixed in every experimental session and was arranged in a manner designed to build aviator proficiency. The order of conditions was determined logically by the author and, judging from the performance data, allowed the aviators to reach the decelerating part of the learning curve before the end of the experiment.

Unfortunately, the manipulation of workload (through the addition of tasks to conditions) is confounded with presentation order. Thus, other factors such as fatigue or a simple order effect may be argued to account for the variance in the data. Randomization of the conditions on the final day could have eliminated the confound at the expense of reducing skill acquisition. Indeed, some instances of catastrophic failure may have been observed if aviators had started the session with the four-task condition. The fact that the TAWL predictions and most of the measures produce rank orderings of the conditions that differ from their presentation order supports the claim that the intended manipulation of workload, and not another variable, accounted for the observations made here.

Suggestions for Further Research

As mentioned above, the current research can be interpreted as support for the validity of the TAWL methodology. It provides a valuable first step in the full validation of the methodology. Of course, if the methodology had not produced valid models in this task environment, then further expense for validation could be saved. The results, however, justify more extensive research.

To date, two research plans have been developed that define the conduct of research to validate TAWL models. Aldrich and Szabo (1986) present a detailed plan for conducting research required to validate the one-crewmember light helicopter, experimental (LHX) workload prediction model. The plan explains the background, defines the problems, and presents the technical objectives to be achieved by conducting the validation research. It includes a review of critical issues with reference to the workload literature. The conduct of the research was planned for an advanced research simulator and includes reference to 174 objective measures of performance that could be used as criteria for validating the model.

The plan calls for two phases of simulator research. In the first phase, segments of the helicopter's mission would be simulated while subjective and performance measures of workload are collected. The presentation of segments would be repeated and randomized. In the second phase, full-mission simulations would be conducted using the same measurement protocol. The plan calls for the refinement of the model whenever sufficient information warrants.

Hamilton (1990) developed a research plan to test the validity of the AH-64A workload prediction model. He modeled the plan after that of Aldrich and Szabo's (1986); however, he designed it to be conducted in the AH-64A combat mission simulator (CMS). Because the CMS is a training simulator rather than a research simulator, very few objective measures of aircrew performance are available from the system. Therefore, the plan describes the development of a performance rating scale intended to be used as a criterion for TAWL validation. The scale measures the tactical performance of the crew on the battlefield. Hamilton (1990) also calls for two phases of research with part- and full-mission simulation.

Recently, the U.S. Army Research Institute (ARI) has acquired a unique research simulation system. The simulator, previously designated the simulator complexity testbed, is now referred to as the simulator training research advanced testbed for aviation (STRATA). The testbed simulates an AH-64A helicopter and was designed to support research in training and simulator design.

Three aspects of the simulator combine to make it the logical site for further TAWL validation research. First, STRATA simulates an AH-64A helicopter for which a TAWL workload prediction model and validation plan already exist.

Second, ARI equipped the simulator with a sophisticated data recording and analysis (DRA) system that can automatically capture and format an enormous number of measures of performance (e.g., airspeed, altitude, ground track). The DRA can also measure predefined sequences of cockpit actions of interest. For example, the researcher can program the DRA to report the average time between enabling and firing the weapon systems computed over the entire mission or to report the average variation in hover position measured only between the enabling and firing of the missile system. These examples are not suggested as measures of performance. They demonstrate that the DRA system offers unique possibilities in the selection of aircrew performance measures.

The final aspect of STRATA that makes it an excellent site for the full validation of the TAWL methodology is in the richness of its tactical environment. The characteristics of both the aircraft and tactical environment influence a pilot's workload. The tactical environment of most simulators is quite impoverished. ARI has equipped the STRATA simulator with an advanced system for the simulation of the tactical environment called the interactive tactical environment management system (ITEMS). The system controls the tactical environment and allows for the control of the simulation down to the physics of individual bullets. ITEMS presents the aviators with unprecedented tactical realism through the use of artificially intelligent opponents and allies. This intelligent simulation is afforded to each of the defined levels of organization (e.g., crew, platoon, company, battalion, etc.). Besides the artificially intelligent players that the system can generate, several auxiliary stations exist that allow SMEs to "take over" any of the players in the simulation to add to the realism.

In summary, two well defined plans exist that define explicit conduct of TAWL validation research. The STRATA facility is recommended as the site for the research because of several unique advantages of the system. Finally, the expense required to update the plan proposed in Hamilton (1990) from the CMS to the STRATA should be minimal.

Summary and Conclusions

This research achieved three objectives. First, it tested the validity of the TAWL methodology by assessing the relationships between subjective, performance, and physiological measures of workload and the predictions generated by two TAWL models. Second, it examined the differences between two models developed by different analysts and evaluated the implications to the validity of the models. Finally, it refined and extended the mechanisms used in the methodology to link workload predictions and performance. The conduct of the research resulted in the following individual findings:

- Seven conditions that vary in workload can be constructed from combinations of four tracking and dichotic listening tasks.

- Two qualified analysts were able to use the TAWL methodology to construct workload prediction models and generate predictions of the workload in each of the conditions.

- Three types of differences (structural, rating, and opinion) were found in the construction of the two models.

- The models' predictions differed but were highly correlated ($r = .99$).

- The NASA TLX ratings of subjective workload were well organized and sensitive to the task loading differences between the conditions.

- The RMS measure of tracking performance consistently increased as a function of condition and had reached a plateau by the last practice session.

- Two measures of dichotic listening performance (omission and exact match) were well organized and sensitive to task condition.

- The heart rate variability measures were not well organized or sensitive to task condition.

- Three measures showed sufficient sensitivity to the across-condition workload manipulation to be used as criteria for TAWL validity.

- The correlations among the 3 criterion measures and the average of TAWL component predictions were high ($.89 < r < .99$) for both models.

These findings led to the following conclusions:

- The TAWL methodology has potential for generating valid predictions of operator workload.

- Although differences were observed between the models generated by different analysts, the differences did not reduce the validity of either model.

- Although the relationship between operator workload and performance is complex (even small amounts of workload can sometimes degrade performance), the TAWL methodology provides an excellent description of the aspects of the task environment known to affect operator performance.

- The validation of the methodology should continue with the validation of a TAWL model of full complexity in the STRATA simulator facility.

References

Aasman, J., Mulder G., & Mulder, L.J.M. (1987). Operator effort and the measurement of heart rate variability. Human Factors, 29(2), 161-170.

Aldrich, T. B., Craddock, W., & McCracken, J. H. (1984). A computer analysis to predict crew workload during LHX scout-attack missions (Technical Report No. ASI479-054-84[B], Vols. I, II, III). Fort Rucker, AL: Anacapa Sciences, Inc.

Aldrich, T. B., & Szabo, S. M. (1986). Validation of the LHX one-crewmember workload prediction model (Technical Memorandum ASI678-202-86[B]). Fort Rucker, AL: Anacapa Sciences, Inc.

Aldrich, T. B., Szabo, S. M., & Craddock, W. (1986). A computer analysis of LHX automation options and their effect on predicted workload (Draft Technical Report ASI479-063-85[B]). Fort Rucker, AL: Anacapa Sciences, Inc.

Bierbaum, C. R., & Aldrich, T. B. (1989). Task analysis of the CH-47D mission and decision rules for developing a CH-47D workload prediction model. Volume I: Summary Report (Research Product 90-10a); Volume II: Appendixes F through I (Research Product 90-10b). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A221 969)

Bierbaum, C. R., & Hamilton, D. B. (1990). Task analysis and workload prediction model of the MH-60K mission and a comparison with UH-60A workload predictions. Volume I: Summary report (Research Report 1576)(AD A241 204); Volume II: Appendixes A through G (RN 91-01) (AD A228 947); Volume III: Appendixes H through N (RN 91-02) (AD A229 408) . Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Bierbaum, C. R., & Hamilton, D. B. (1991). Task analysis and workload prediction for the MH-47E and a comparison with CH-47D workload predictions. Volume I: Summary report (Research Report 1584) (AD A237 500); Volume II: Appendixes A through N (RN 91-35) (AD A235 249). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Bierbaum, C. R., Szabo, S. M., & Aldrich, T. B. (1989). Task analysis of the UH-60 mission and decision rules for developing a UH-60 workload prediction model. Volume I: Summary Report (Research Product 89-08). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A210 763)

Bryden, M. P. (1969). Binaural competition and division of attention as determinants of the laterality effect in dichotic listening. Canadian Journal of Psychology, 23, 101-113.

Byers, J. C., Bittner, A. C., & Hill, S. G. (1989). Traditional and raw Task Load Index (TLX) correlations: Are pared comparisons necessary? In Advances in Industrial Ergonomics and Safety, Vol. 1. London: Taylor and Francis.

Geisser, S., & Greenhouse, S. W. (1958). An extension of Box's results on the use of the F distribution in multivariate analysis. Annals of Mathematical Statistics, 29, 885-891.

Gopher, D. (1980). On the training of time-sharing skills: An attention viewpoint. In Proceedings of the Human Factors Society, 24th Annual Meeting. Santa Monica, CA: Human Factors Society.

Gopher, D. (1982). A selective attention test as a predictor of success in flight training. Human Factors, 24(2), 173-183.

Gopher, D., Brikner, M., & Navon, D. (1980). Different difficulty manipulations interact differently with task emphasis: Evidence for multiple resources (Technical Report No. AFPSR-77-31313). Haifa, Israel: Technion.

Gopher, D. & Kahneman, D. (1971). Individual differences in attention and the prediction of flight criteria. Perceptual and Motor Skills, 33, 1335-1342.

Griffin, G. R., & McBride, D. K. (1986). Multitask performance: Predicting success in naval aviation primary flight training (Technical Report No. NAMRL-1316). Pensacola, FL: Naval Aerospace Medical Research Laboratory.

Hamilton, D. B. (1993). Validation of the AH-64A workload model (Working Paper No. ARIARDA/ASI 92-04). Fort Rucker, AL: Army Reseach Institute Aviation Research and Development Activity.

Hamilton, D. B., & Bierbaum, C. R. (1992). Operator workload predictions for the revised AH-64A workload prediction model (ARI Research Note 92-61). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A254 198)

Hamilton, D. B., Bierbaum, C. R., & Fulford, L. A. (1991). Task analysis/workload (TAWL) user's guide - version 4.0 (Research Product 91-11). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A241 861)

Iavecchia, H. P., Linton, P. M., Bittner, A. C., Jr., & Byers, J. C. (1989). Operator workload in the UH-60 Black Hawk: Crew results vs. TAWL model predictions. In Proceedings of the Human Factors Society, 33rd Annual Meeting. Santa Monica, CA: Human Factors Society.

Intano, G. P., Howse, W. R., & Lofaro, R. J. (1991a). The selection of an experimental test battery for aviator cognitive, psychomotor abilities and personal traits (Research Note 91-21). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A231 887)

Intano, G. P., Howse, W. R., & Lofaro, R. J. (1991b). Initial validation of the Army aviator classification process (Research Note 91-38). Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences. (AD A235 768)

Intano, G. P., & Lofaro, R. J. (1989). Army aviator classification by aircraft type. Human Factors Society Bulletin, 32(1), 2-4.

Intano, G. P., & Lofaro, R. J. (1990). Initial validation of the Army aviator candidate classification algorithm. Human Factors Society Bulletin, 33(4), 1-3.

Kimura, D. (1967). Functional asymmetry of the brain in dichotic listening. Cortex, 3, 163-178.

McCracken, J. H., & Aldrich, T. B. (1984). Analyses of selected LHX mission functions: Implications for operator workload and system automation goals (Technical Note ASI479-024-84[B]). Fort Rucker, AL: Anacapa Sciences, Inc.

NASA Task Load Index (TLX): Paper-and-Pencil Version. (1986). Moffett Field, CA: NASA-Ames Research Center, Aerospace Human Factors Research Division.

Schneider, W., & Fisk, A. D. (1980). Dual task automatic and controlled processing in visual search (HARL Technical Report No. 8002). Champagne, IL: University of Illinois, Department of Psychology.

Szabo, S. M., & Bierbaum, C. R. (1986). A comprehensive task analysis of the AH-64 mission with workload estimates and preliminary decision rules for developing an AH-64 workload prediction model (Technical Report No. ASI678-204-86[B], Vols. I, II, III, and IV). Fort Rucker, AL: Anacapa Sciences, Inc.

Treisman, A. M., & Geffen, G. (1968). Selective attention and cerebral dominance in perceiving and responding to speech messages. Quarterly Journal of Experimental Psychology, 20, 139-151.

Van Dellen, H. H., Aasman, J., Mulder, L. S. M., & Mulder, G. (1985). Time domain versus frequency domain measures of heart rate variability. In J. F. Orlebeke, G. Mulder, & L. J. P. van Doormen (Eds.), Psychophysiology of Cardiovascular Control: Models, Methods and Data (pp. 353-374). New York: Plenum.

Wickens, C. D., & Gopher, D. (1977). Control theory measures of tracking as indices of attention allocation strategies. Human Factors, 19, 349-365.

Wickens, C. D. (1984). Engineering psychology and human performance. Columbus, OH: Merrill.

APPENDIX A

DEMOGRAPHIC SURVEY

TAWL WORKLOAD PREDICTION METHODOLOGY VALIDATION RESEARCH

AVIATOR DEMOGRAPHIC SURVEY

INSTRUCTIONS: This survey consists of questions that provide information about your personal background and experience. Answer each item that applies to you by checking in the appropriate bracket [√] or by printing the required information in the space provided. When answering items about flight hours, you may refer to records, if available, or you may estimate the flight hours as closely as possible. Your responses will be used for research purposes only.

1. Name: ______________________________________
   Last First Middle

2. Social Security Number: ___ ___ ___ - ___ ___ - ___ ___ ___ ___

3. Today's Date: ______________________________
   Month / Day / Year

4. What is your age ?
   ________ Years

5. What is your current rank?
   [ ] WO1 [ ] 2LT
   [ ] CW2 [ ] 1LT
   [ ] CW3 [ ] CPT
   [ ] CW4 [ ] MAJ
   [ ] LTC
   [ ] COL

6. To which unit are you assigned?
   Unit: ______________________________

7. How long have you been assigned to your present Company?
   ________ years and _______ months

8. Do you anticipate reassignment in the next 6 months?
   [ ] Yes
   [ ] No

   If yes, give expected date and location of reassignment________________
   ____

9. Currently, what is your primary duty position in the unit?

____

10. What additional duties do you perform in your unit?

____

____

____

11. How long have you been on active duty military service?
_______ years and _______ months of active service

12. How long has it been since you graduated from initial Army flight training?
_______ years and _______ months

13. How long has it been since you graduated from the AH-64 AQC?
_______ years and _______ months

14. Were you an IERW turnaround student in the AH-64 AQC?
[ ] Yes
[ ] No

If no, what was your primary aircraft before entering the AH-64 AQC?

____

15. Indicate the total number of flight hours you have logged in each of the following aircraft. Also, check [√] the highest duty category you have held in each aircraft.

a. **Military Rotary Wing**

			PI	PC	UT	IP	SI	IE
AH-64	_______	hours	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]
AH-1	_______	hours	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]
OH-58	_______	hours	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]
UH-1	_______	hours	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]
Other	_______	hours	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]

(Specify other aircraft)______________________________

b. **Military Fixed Wing**

U-21	_______	hours	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]
C-12	_______	hours	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]
OV-1	_______	hours	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]
Other	_______	hours	[ ]	[ ]	[ ]	[ ]	[ ]	[ ]

(Specify other aircraft)______________________________

16. How many flight hours have you logged in each seat of the AH-64?
Front Seat __________ hours
Back Seat __________ hours

17. How many flight hours have you logged in each seat of the AH-64 CMS?
Front Seat __________ hours
Back Seat __________ hours

18. How many flight hours have you logged in each seat of the AH-64 CWEPT?
Front Seat __________ hours
Back Seat __________ hours

19. What is your current crew station designation?
[ ] AH-64 front seat
[ ] AH-64 back seat
[ ] Both seats (explain) ________________________________

20. What is your current Readiness Level?
[ ] RL1
[ ] RL2
[ ] RL3

21. Have you been assigned to a fixed crewmate?
[ ] Yes
[ ] No

22. If you are a member of a fixed crew, how many hours has your crew trained together?
__________ flight hours
__________ CMS hours

A P P E N D I X   B

NASA Task Load Index (TLX) Instructions

Workload Prediction Validation Research

NASA-Task Load Index Rating Scales

Orientation: During this research, we are interested in assessing your performance and your experiences during individual segments of a tracking and listening test battery. Below, the technique that will be used to examine your experiences is described. The technique was developed by NASA and is called the Task Load Index or TLX for short. In the most general sense, we want to examine your "workload" experience. Workload is a difficult concept to define precisely, but a simple one to understand. The factors that influence your experience of workload may come from the task itself, your feelings about your own performance, how much effort you put in, or the stress and frustration you feel. The workload contributed by different tasks may change as you get more familiar with a task, perform easier or harder versions of the task, or move from one task to another. Physical workload is relatively easy to conceptualize and evaluate. However, mental workload is more subjective, making it more difficult to measure.

Because workload is something that is an individual experience, effective "rulers" of workload are difficult to define. Previous research has identified at least six factors that may contribute to workload. Because workload may be caused by any of these factors, we would like you to evaluate each of them rather than lumping them into an overall evaluation of workload. The set of six rating scales described on the following page was developed to evaluate your experiences during different task segments. Please read the descriptions of the scales carefully. If you have a question about any of the scales in the table, please ask me about it. It is extremely important that they be clear to you. You may keep the descriptions with you for reference during the experiment.

Instructions: For each task segment that you perform, you will be given a sheet of six rating scales similar to the one pictured above. After each segment, you will evaluate the segment by putting a mark on each of the six scales at the point which matches your experience. Each scale has two single word descriptors that anchor each endpoint. Please consider your responses carefully in distinguishing among the different task segments. Consider each scale individually. Your ratings will play an important role in the evaluation being conducted. Thus, your active participation is essential to the success of this experiment and is greatly appreciated.

Workload Prediction Validation Research

NASA-Task Load Index
Rating Scales

Rating Scale Definitions		
Title	*Endpoints*	Descriptions
MENTAL DEMAND	*Low / High*	How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, forgiving or exacting?
PHYSICAL DEMAND	*Low / High*	How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
TEMPORAL DEMAND	*Low / High*	How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
PERFORMANCE	*Poor / Good*	How successful do you think you were in accomplishing the goals of the task set by yourself? How satisfied were you with your performance in accomplishing these goals?
EFFORT	*Low / High*	How hard did you have to work (mentally or physically) to accomplish your level of performance?
FRUSTRATION LEVEL	*Low / High*	How secure, gratified, content, relaxed and complacent versus insecure, discouraged, irritated, stressed and annoyed did you feel during the task?

Workload Prediction Validation Research

NASA-Task Load Index
Sources of Workload

Orientation: Throughout this research, six rating scales were used to assess your experiences in the different task segments. Scales of this sort are extremely useful, but their utility suffers from the tendency people have to interpret them in individual ways. For example, some people feel that mental or temporal demands are the essential aspects of workload regardless of the effort they expended on a given task or the level of performance they achieved. Others feel that if they performed well, the workload must have been low and if they performed badly, it must have been high. Yet others feel that effort or feelings of frustration are the most important factors in workload, and so on. In addition, the factors that create levels of workload differ depending on the task. For example, some tasks might be difficult because they must be completed very quickly. Others may seem easy or hard because of the intensity of mental or physical effort required. Yet others feel difficult because they cannot be performed well, no matter how much effort is expended.

Instructions: The evaluation you are about to perform is a technique that has been developed by NASA to assess the relative importance of the six factors in determining your overall workload. The procedure is simple: You will be presented with a series of pairs of rating scale titles (for example, Effort or Mental Demand) and asked to choose which of the items was more important to ***your*** overall experience of workload in the segments that you performed. Each pair of scale titles will appear on a separate card.

Circle the scale title that represents the more important contributor to workload for the segments you performed in this research.

After you have finished the entire series, we will be able to use the pattern of your choices to create a weighted combination of your segment ratings to generate an overall workload score. Please consider your choices carefully and make them consistent with how you used the rating scales during the task segments you were asked to evaluate. Don't think that there is any correct pattern; we are only interested in your opinions.

If you have any questions, please ask them now. Otherwise, start whenever you are ready. Thank you for your participation.

Task 1. - Stick Only

MENTAL DEMAND

Low High

PHYSICAL DEMAND

Low High

TEMPORAL DEMAND

Low High

PERFORMANCE

Poor Good

EFFORT

Low High

FRUSTRATION

Low High

A P P E N D I X   C

EXPERIMENTAL CONTROL FILE

Appendix C contains the computer file that controlled the experimental sessions of the TAWL workload validation research. The file contains two types of controlling instructions that are placed in the first column of the file. The I instructions indicate that the following material is an instruction screen. The E instructions indicate that the next line in the file contains the specification for the next experimental trial. Experimental conditions were created by selecting the dichotic listening (D), stick (S), pedals (P), or throttle (T) tasks to be included in the trail using the Dn Sn Pn Tn command. Replacing the n with either a 1 or a 0 would activate or deactivate the tasks, respectively.

I

* Welcome to the *
* Task Analysis / Workload Prediction Validation Research *

During this experiment you will be performing 7 different tasks. You will be repeating each task 3 times for a total of 21 trials. Each trial takes 3 minutes. The experiment is entirely self-paced. You determine when you perform the next trial. You may take up to 1 minute between trials and finish in the allotted time. As an aid in keeping pace, an elapsed time clock is provided on each screen.

We are recording your heart rate, respirations, and eye blinks, please notify the experimenter if the adhesive on any of the electrodes fails.

Please check that the RED wire enters the headset on the right.

Remember your concentration and improvement on these tasks is important to the outcome of this experiment. Try your best.

When you are ready for the Task 1 instructions...Press the #1 key

I

* Task 1 - Stick Only *

In this experiment only the center controller (where the cyclic in a helicopter would be) will be active. Obtain a comfortable seating position placing your feet on the pedals and let your left hand rest to your side. The trigger and buttons on the controller have no effect on the computer. Grasp the controller handle with your right hand placing your index finger on the trigger. When the trial begins, use the handle to control the movements of the cross in the central portion of the computer screen. Try to maintain the cross in the center of the screen in line with the cross hairs.

The computer has a built in drift component so you are following a moving target.

When you are ready to begin Task 1 - Trial 1

Press the #1 key!

E
D0 S1 P0 T0
I

Task 1 - Trial 1 completed!

* Relax, take a break for a minute... *

When you are ready to begin Task 1 - Trial 2

Press the #1 key!

E
D0 S1 P0 T0

I

Task 1 - Trial 2 completed!

* Relax, take a break for a minute... *

When you are ready to begin Task 1 - Trial 3

Press the #1 key!

E
DO S1 PO TO
I

Task 1 - Trial 3 completed!

* Take a minute to complete the NASA-TLX rating scale for *   * Task 1 *

When you are ready for the Task 2 instructions...
Press the #1 key

I

* Task 2 - Listening Only *

In this experiment you will be listening to sets of letters and digits spoken over the headphones and entering your responses on the number keypad with your left hand. During each set, letters and digits are presented to each ear simultaneously. YOUR JOB IS TO FOCUS YOUR ATTENTION ON ONE EAR AND TO ENTER ONLY THE DIGITS HEARD THROUGH THAT EAR ON THE KEYPAD. The ear you focus on changes from set to set and will be clearly identified as RIGHT or LEFT before each set of letters and digits begins. Use your index finger to record the digits digits you hear by pressing the key corresponding to the digit. Press the appropriate key as soon as you hear the digit.

Remember, the letter 'O' is NOT the number 'ZERO'
When you are ready to begin Task 2 - Trial 1

Press the #1 key!

E
D1 SO PO TO

I

Task 2 - Trial 1 completed!

* Relax, take a break for a minute... *

When you are ready to begin Task 2 - Trial 2

Press the #1 key!

E
D1 S0 P0 T0
I

Task 2 - Trial 2 completed!

* Relax, take a break for a minute... *

When you are ready to begin Task 2 - Trial 3

Press the #1 key!

E
D1 S0 P0 T0
I

Task 2 - Trial 3 completed!

* Take a minute to complete the NASA-TLX rating scale for Task 2 *

When you are ready for the Task 3 instructions...

Press the #1 key

I

* Task 3 - Listening + Stick *

You will now perform Task 1 (stick) and Task 2 (listening) in dual. Try to do each task equally well. You may experience difficulty, but do your very best. First the stick task will initiate. A few seconds later later, the listening task will start. The listening task will end just before the stick task. Be certain to continue performing the stick task until the video display disappears from the screen.

When you are ready to begin Task 3 - Trial 1

Press the #1 key!

E
D1 S1 P0 T0
I

Task 3 - Trial 1 completed!

* Relax, take a break for a minute... *

When you are ready to begin Task 3 - Trial 2

Press the #1 key!

E
D1 S1 P0 T0
I

Task 3 - Trial 2 completed!

* Relax, take a break for a minute... *

When you are ready to begin Task 3 - Trial 3

Press the #1 key!

E
D1 S1 P0 T0

I

Task 3 - Trial 3 completed!

* Take a minute to complete the NASA-TLX rating scale for *   * Task 3 *

When you are ready for the Task 4 instructions...

Press the #1 key

I

* Task 4 - Stick + Pedals *

In this experiment only the center controller and the pedals will be active. Obtain a comfortable seating position placing your right hand on the center controller. Place your feet on the two foot pedals, so that your feet rest halfway on the pedals and your heels on the floor. These foot pedals will control the direction and speed of the foot pedal cursor. Pressing on the left pedal moves the cursor right. Pressing on the right pedal moves the cursor left. When the trial begins, you will perform the stick tracking task as before. You will also see the foot pedal cursor move across the bottom of the screen. You must keep the foot pedal cursor as close as possible to the vertical bar in the center of the screen. The computer has a built in drift component so you are following moving targets.

When you are ready to begin Task 4 - Trial 1

Press the #1 key!

E
D0 S1 P1 T0
I

Task 4 - Trial 1 completed!

* Relax, take a break for a minute... *

When you are ready to begin Task 4 - Trial 2

Press the #1 key!

E
D0 S1 P1 T0

I

Task 4 - Trial 2 completed!

* Relax, take a break for a minute... *

When you are ready to begin Task 4 - Trial 3

Press the #1 key!

E
D0 S1 P1 T0
I

Task 4 - Trial 3 completed!

* Take a minute to complete the NASA-TLX rating scale for *   * Task 4 *

When you are ready for the Task 5 instructions...

Press the #1 key

I

* Task 5 - Listening + Stick + Pedals *

In this experiment you will be performing the listening task, the stick task, and the pedal task in combination. Try to do each task equally well.

When you are ready to begin Task 5 - Trial 1

Press the #1 key!

E
D1 S1 P1 T0

I

Task 5 - Trial 1 completed!

* Relax, take a break for a minute... *

When you are ready to begin Task 5 - Trial 2

Press the #1 key!

E
D1 S1 P1 T0
I

Task 5 - Trial 2 completed!

* Relax, take a break for a minute... *

When you are ready to begin Task 5 - Trial 3

Press the #1 key!

E
D1 S1 P1 T0
I

Task 5 - Trial 3 completed!

* Take a minute to complete the NASA-TLX rating scale for *   * Task 5 *

When you are ready for the Task 6 instructions...

Press the #1 key

I

* Task 6 - Stick + Pedals + Throttle *

In this experiment the center controller, the pedals, and the right arm controller will be active. You will perform the stick tracking and foot pedal tasks as before. Position your right hand and feet as in the previous tests. When the trial begins, you will also see the throttle cursor move along the left edge of the screen. The throttle is controlled by the handle located to your left. Take hold of the throttle by closing your LEFT hand around the handle so that the handle is inside your left fist and your index finger is on the trigger. Pressing the stick forward moves the cursor up; pulling it toward you moves it down. You must keep the throttle cursor as close as possible to the horizontal bar in the center of the screen.

When you are ready to begin Task 6 - Trial 1

Press the #1 key!

E
DO S1 P1 T1
I

Task 6 - Trial 1 completed!

* Relax, take a break for a minute... *

When you are ready to begin Task 6 - Trial 2

Press the #1 key!

E
DO S1 P1 T1
I

Task 6 - Trial 2 completed!

* Relax, take a break for a minute... *

When you are ready to begin Task 6 - Trial 3

Press the #1 key!

E
DO S1 P1 T1

I

Task 6 - Trial 3 completed!

* Take a minute to complete the NASA-TLX rating scale for *   * Task 6 *

When you are ready for the Task 7 instructions...

Press the #1 key

I

* Task 7 - Listening + Stick + Pedals + Throttle *

Its all on now! In this experiment the center controller, left controller, and the pedals will active. In addition the listening task will be active. It is almost impossible to control the throttle and enter numbers on the keypad at the same time unless you use your forearm or use some other innovation. This is a difficult task. Try to do your best.

When you are ready to begin Task 7 - Trial 1

Press the #1 key!

E
D1 S1 P1 T1
I

Task 7 - Trial 1 completed!

* Relax, take a break for a minute... *

When you are ready to begin Task 7 - Trial 2

Press the #1 key!

E
D1 S1 P1 T1

I

Task 7 - Trial 2 completed!

* Relax, take a break for a minute... *

When you are ready to begin Task 7 - Trial 3

Press the #1 key!

E
D1 S1 P1 T1
I

Task 7 - Trial 3 completed!

* Take a minute to complete the NASA-TLX rating scale for * * Task 7 *

When you have completed the NASA-TLX rating...

Press the #1 key

I

* Thank you for your participation in the * * * Task Analysis / Workload Prediction Validation Research *

That's all folks!

.

A P P E N D I X D

TAWL Model 1

Preliminary TAWL Validation Workload Prediction Model - CRB

TASK DATABASE

Task	COG	PSY	VIS	AUD	KIN
1. CONTROL STICK	1.20	2.60	0.00	0.00	1.00
2. CONTROL THROTTLE	1.20	2.60	0.00	0.00	1.00
3. RECEIVE RANDOM NUMBERS	5.30	0.00	0.00	4.90	0.00
4. MONITOR AUDIO	1.00	0.00	0.00	1.00	0.00
5. ENTER RANDOM NUMBERS	1.20	2.20	3.70	0.00	1.00
6. CONTROL PEDALS	1.20	2.60	0.00	0.00	1.00
7. MONITOR CRT	1.20	0.00	5.40	0.00	0.00

Preliminary TAWL Validation Workload Prediction Model - CRB

Function Decision Rules

Function 1: CONTROL VERTICAL MOVEMENT

Discrete Fixed Tasks

Entry #	Task #	Task Name	Crewmember	Start	Duration
1	2	CONTROL THROTTLE	PILOT (P)	0.0	-0.5

Preliminary TAWL Validation Workload Prediction Model - CRB

Function Decision Rules

Function 2: CONTROL HORIZONTAL MOVEMENT

Discrete Fixed Tasks

Entry #	Task #	Task Name	Crewmember	Start	Duration
1	6	CONTROL PEDALS	PILOT (P)	0.0	-0.5

Preliminary TAWL Validation Workload Prediction Model - CRB

Function Decision Rules

Function 3: CONTROL RETICLE MOVEMENT

Discrete Fixed Tasks

Entry #	Task #	Task Name	Crewmember	Start	Duration
1	1	CONTROL STICK	PILOT (P)	0.0	-0.5

Preliminary TAWL Validation Workload Prediction Model - CRB

Function Decision Rules

Function 4: ENTERING NUMBERS

Discrete Fixed Tasks

Entry #	Task #	Task Name	Crewmember	Start	Duration
1	5	ENTER RANDOM NUMBERS	PILOT (P)	0.0	1.0

Preliminary TAWL Validation Workload Prediction Model - CRB

Function Decision Rules

Function 5: MONITOR VISUAL

Discrete Fixed Tasks

Entry #	Task #	Task Name	Crewmember	Start	Duration
1	7	MONITOR CRT	PILOT (P)	0.0	-0.5

Preliminary TAWL Validation Workload Prediction Model - CRB

Function Decision Rules

Function 6: MONITOR AUDIO

Discrete Fixed Tasks

Entry #	Task #	Task Name	Crewmember	Start	Duration
1	4	MONITOR AUDIO	PILOT (P)	0.0	~0.5

Preliminary TAWL Validation Workload Prediction Model - CRB

Function Decision Rules

Function 7: RECEIVE NUMBERS

Discrete Fixed Tasks

Entry #	Task #	Task Name	Crewmember	Start	Duration
1	3	RECEIVE RANDOM NUMBERS	PILOT (P)	0.0	12.5
2	4	MONITOR AUDIO	PILOT (P)	12.5	1.5
3	3	RECEIVE RANDOM NUMBERS	PILOT (P)	14.0	4.0
4	4	MONITOR AUDIO	PILOT (P)	18.0	2.5

Preliminary TAWL Validation Workload Prediction Model - CRB

Segment Decision Rules

Segment 1: DL ONLY

Discrete Fixed Functions

#	Function Name	Start	Duration	Interrupted By
6	MONITOR AUDIO	0.0	8.0	7
7	RECEIVE NUMBERS	8.0	20.5	
7	RECEIVE NUMBERS	28.5	20.5	
7	RECEIVE NUMBERS	49.0	20.5	
7	RECEIVE NUMBERS	69.5	20.5	
7	RECEIVE NUMBERS	90.0	20.5	
7	RECEIVE NUMBERS	110.5	20.5	
7	RECEIVE NUMBERS	131.0	20.5	
7	RECEIVE NUMBERS	151.5	20.5	
6	MONITOR AUDIO	172.0	-8.0	

Discrete Random Functions

#	Function Name	Duration	Start	Finish	Times	Interrupted By
4	ENTERING NUMBERS	1.5	10.0	20.0	5	
4	ENTERING NUMBERS	1.5	21.0	28.5	4	
4	ENTERING NUMBERS	1.5	30.5	40.5	5	
4	ENTERING NUMBERS	1.4	41.5	49.0	4	
4	ENTERING NUMBERS	1.5	51.0	61.0	5	
4	ENTERING NUMBERS	1.5	62.0	69.5	4	
4	ENTERING NUMBERS	1.5	71.5	81.5	5	
4	ENTERING NUMBERS	1.5	82.5	90.0	4	
4	ENTERING NUMBERS	1.5	92.0	102.0	5	
4	ENTERING NUMBERS	1.5	103.0	110.5	4	
4	ENTERING NUMBERS	1.5	112.5	122.5	5	
4	ENTERING NUMBERS	1.5	123.5	131.0	4	
4	ENTERING NUMBERS	1.5	133.0	143.0	5	
4	ENTERING NUMBERS	1.5	144.0	151.5	4	
4	ENTERING NUMBERS	1.5	153.5	163.5	5	
4	ENTERING NUMBERS	1.5	164.5	180.0	4	

Preliminary TAWL Validation Workload Prediction Model - CRB

Segment Decision Rules

Segment 2: STICK ONLY

Continuous Fixed Functions

#	Function Name	Start	Duration	Interrupted By
3	CONTROL RETICLE MOVEMENT	0.0	180.0	
5	MONITOR VISUAL	0.0	180.0	

Preliminary TAWL Validation Workload Prediction Model - CRB

Segment Decision Rules

Segment 3: DL AND STICK

Discrete Fixed Functions

#	Function Name	Start	Duration	Interrupted By
6	MONITOR AUDIO	0.0	8.0	7
7	RECEIVE NUMBERS	8.0	20.5	
7	RECEIVE NUMBERS	28.5	20.5	
7	RECEIVE NUMBERS	49.0	20.5	
7	RECEIVE NUMBERS	69.5	20.5	
7	RECEIVE NUMBERS	90.0	20.5	
7	RECEIVE NUMBERS	110.5	20.5	
7	RECEIVE NUMBERS	131.0	20.5	
7	RECEIVE NUMBERS	151.5	20.5	
6	MONITOR AUDIO	172.0	-8.0	

Continuous Fixed Functions

#	Function Name	Start	Duration	Interrupted By
3	CONTROL RETICLE MOVEMENT	0.0	-180.0	
5	MONITOR VISUAL	0.0	180.0	4

Discrete Random Functions

#	Function Name	Duration	Start	Finish	Times	Interrupted By
4	ENTERING NUMBERS	1.0	10.0	20.0	5	
4	ENTERING NUMBERS	1.0	21.0	28.5	4	
4	ENTERING NUMBERS	1.0	30.5	40.5	5	
4	ENTERING NUMBERS	1.0	41.5	49.0	4	
4	ENTERING NUMBERS	1.0	51.0	61.0	5	
4	ENTERING NUMBERS	1.0	62.0	69.5	4	
4	ENTERING NUMBERS	1.0	71.5	81.5	5	
4	ENTERING NUMBERS	1.0	82.5	90.0	4	
4	ENTERING NUMBERS	1.0	92.0	102.0	5	
4	ENTERING NUMBERS	1.0	103.0	110.5	4	
4	ENTERING NUMBERS	1.0	112.5	122.5	5	
4	ENTERING NUMBERS	1.0	123.5	131.0	4	
4	ENTERING NUMBERS	1.0	133.0	143.0	5	
4	ENTERING NUMBERS	1.0	144.0	151.5	4	
4	ENTERING NUMBERS	1.0	153.5	163.5	5	
4	ENTERING NUMBERS	1.0	164.5	180.0	4	

Preliminary TAWL Validation Workload Prediction Model - CRB

Segment Decision Rules

Segment 4: STICK AND PEDAL

Continuous Fixed Functions

#	Function Name	Start	Duration	Interrupted By
1	CONTROL VERTICAL MOVEMENT	0.0	180.0	
3	CONTROL RETICLE MOVEMENT	0.0	180.0	
5	MONITOR VISUAL	0.0	180.0	

Preliminary TAWL Validation Workload Prediction Model - CRB

Segment Decision Rules

Segment 5: DL, STICK AND PEDAL

Discrete Fixed Functions

#	Function Name	Start	Duration	Interrupted By
6	MONITOR AUDIO	0.0	8.0	7
7	RECEIVE NUMBERS	8.0	20.5	
7	RECEIVE NUMBERS	28.5	20.5	
7	RECEIVE NUMBERS	49.0	20.5	
7	RECEIVE NUMBERS	69.5	20.5	
7	RECEIVE NUMBERS	90.0	20.5	
7	RECEIVE NUMBERS	110.5	20.5	
7	RECEIVE NUMBERS	131.0	20.5	
7	RECEIVE NUMBERS	151.5	20.5	
6	MONITOR AUDIO	172.0	8.0	

Continuous Fixed Functions

#	Function Name	Start	Duration	Interrupted By
1	CONTROL VERTICAL MOVEMENT	0.0	180.0	4
3	CONTROL RETICLE MOVEMENT	0.0	-180.0	
5	MONITOR VISUAL	0.0	180.0	4

Discrete Random Functions

#	Function Name	Duration	Start	Finish	Times	Interrupted By
4	ENTERING NUMBERS	1.0	10.0	20.0	5	
4	ENTERING NUMBERS	1.0	21.0	28.5	4	
4	ENTERING NUMBERS	1.0	30.5	40.5	5	
4	ENTERING NUMBERS	1.0	41.5	49.0	4	
4	ENTERING NUMBERS	1.0	51.0	61.0	5	
4	ENTERING NUMBERS	1.0	62.0	69.5	4	
4	ENTERING NUMBERS	1.0	71.5	81.5	5	
4	ENTERING NUMBERS	1.0	82.5	90.0	4	
4	ENTERING NUMBERS	1.0	92.0	102.0	5	
4	ENTERING NUMBERS	1.0	103.0	110.5	4	
4	ENTERING NUMBERS	1.0	112.5	122.5	5	
4	ENTERING NUMBERS	1.0	123.5	131.0	4	
4	ENTERING NUMBERS	1.0	133.0	143.0	5	
4	ENTERING NUMBERS	1.0	144.0	151.5	4	
4	ENTERING NUMBERS	1.0	153.5	163.5	5	
4	ENTERING NUMBERS	1.0	164.5	180.0	4	

Preliminary TAWL Validation Workload Prediction Model - CRB

Segment Decision Rules

Segment 6: STICK, PEDAL AND THROTTLE

Continuous Fixed Functions

#	Function Name	Start	Duration	Interrupted By
1	CONTROL VERTICAL MOVEMENT	0.0	180.0	
2	CONTROL HORIZONTAL MOVEMENT	0.0	180.0	
3	CONTROL RETICLE MOVEMENT	0.0	180.0	
5	MONITOR VISUAL	0.0	180.0	

Preliminary TAWL Validation Workload Prediction Model - CRB

Segment Decision Rules

Segment 7: DL, STICK, PEDAL AND THROTTLE

Discrete Fixed Functions

#	Function Name	Start	Duration	Interrupted By
6	MONITOR AUDIO	0.0	8.0	7
7	RECEIVE NUMBERS	8.0	20.5	
7	RECEIVE NUMBERS	28.5	20.5	
7	RECEIVE NUMBERS	49.0	20.5	
7	RECEIVE NUMBERS	69.5	20.5	
7	RECEIVE NUMBERS	90.0	20.5	
7	RECEIVE NUMBERS	110.5	20.5	
7	RECEIVE NUMBERS	131.0	20.5	
7	RECEIVE NUMBERS	151.5	20.5	
6	MONITOR AUDIO	172.0	-8.0	

Continuous Fixed Functions

#	Function Name	Start	Duration	Interrupted By
1	CONTROL VERTICAL MOVEMENT	0.0	180.0	4
2	CONTROL HORIZONTAL MOVEMENT	0.0	180.0	
3	CONTROL RETICLE MOVEMENT	0.0	-180.0	
5	MONITOR VISUAL	0.0	180.0	4

Discrete Random Functions

#	Function Name	Duration	Start	Finish	Times	Interrupted By
4	ENTERING NUMBERS	1.0	10.0	20.0	5	
4	ENTERING NUMBERS	1.0	21.0	28.5	4	
4	ENTERING NUMBERS	1.0	30.5	40.5	5	
4	ENTERING NUMBERS	1.0	41.5	49.0	4	
4	ENTERING NUMBERS	1.0	51.0	61.0	5	
4	ENTERING NUMBERS	1.0	62.0	69.5	4	
4	ENTERING NUMBERS	1.0	71.5	81.5	5	
4	ENTERING NUMBERS	1.0	82.5	90.0	4	
4	ENTERING NUMBERS	1.0	92.0	102.0	5	
4	ENTERING NUMBERS	1.0	103.0	110.5	4	
4	ENTERING NUMBERS	1.0	112.5	122.5	5	
4	ENTERING NUMBERS	1.0	123.5	131.0	4	
4	ENTERING NUMBERS	1.0	133.0	143.0	5	
4	ENTERING NUMBERS	1.0	144.0	151.5	4	
4	ENTERING NUMBERS	1.0	153.5	163.5	5	
4	ENTERING NUMBERS	1.0	164.5	180.0	4	

A P P E N D I X E

TAWL Model 2

Preliminary TAWL Validation Workload Prediction Model - DBH

TASK DATABASE

Task	Aud	Vis	Cog	Psy
1. Control Vertical Position	0.00	0.00	1.00	2.60
2. Control Horizontal Position	0.00	0.00	1.00	2.60
3. Monitor Audio	1.00	0.00	0.00	0.00
4. Detect Numbers	4.90	0.00	4.60	0.00
5. Enter Numbers	0.00	5.00	1.20	2.20
6. Monitor Screen	0.00	5.40	0.00	0.00

Preliminary TAWL Validation Workload Prediction Model - DBH

Function Decision Rules

Function 1: Stick

Continuous Random Tasks

Start : 0.0
Finish: -0.5
Crewmember: AVIATOR (A)

Entry #	Task #	Task Name	Duration
1	1	Control Vertical Position	0.5
2	2	Control Horizontal Position	0.5

Preliminary TAWL Validation Workload Prediction Model - DBH

Function Decision Rules

Function 2: Pedals

Continuous Fixed Tasks

Entry #	Task #	Task Name	Crewmember	Start	Duration
1	2	Control Horizontal Position	AVIATOR	0.0	-0.5

Preliminary TAWL Validation Workload Prediction Model - DBH

Function Decision Rules

Function 3: Throttle

Continuous Fixed Tasks

Entry #	Task #	Task Name	Crewmember	Start	Duration
1	1	Control Vertical Position	AVIATOR (A)	0.0	-0.5

Preliminary TAWL Validation Workload Prediction Model - DBH

Function Decision Rules

Function 4: Detect Numbers

Discrete Fixed Tasks

Entry #	Task #	Task Name	Crewmember	Start	Duration
1	3	Monitor Audio	AVIATOR (A)	0.0	1.5
2	4	Detect Numbers	AVIATOR (A)	1.5	11.0
3	3	Monitor Audio	AVIATOR (A)	12.5	1.5
4	4	Detect Numbers	AVIATOR (A)	14.0	4.0

Preliminary TAWL Validation Workload Prediction Model - DBH

Function Decision Rules

Function 5: Monitor Audio

Continuous Fixed Tasks

Entry #	Task #	Task Name	Crewmember	Start	Duration
1	3	Monitor Audio	AVIATOR (A)	0.0	-0.5

Preliminary TAWL Validation Workload Prediction Model - DBH

Function Decision Rules

Function 6: Monitor Screen

Continuous Fixed Tasks

Entry #	Task #	Task Name	Crewmember	Start	Duration
1	6	Monitor Screen	AVIATOR (A)	0.0	-0.5

Preliminary TAWL Validation Workload Prediction Model - DBH

Function Decision Rules

Function 7: Enter Number

Discrete Fixed Tasks

Entry #	Task #	Task Name	Crewmember	Start	Duration
1	5	Enter Numbers	AVIATOR (A)	0.0	0.5

Preliminary TAWL Validation Workload Prediction Model - DBH

Function Decision Rules

Function 8: Enter Numbers

Discrete Fixed Tasks

Entry #	Task #	Task Name	Crewmember	Start	Duration
1	5	Enter Numbers	AVIATOR (A)	0.0	1.0
2	5	Enter Numbers	AVIATOR (A)	1.0	1.0
3	5	Enter Numbers	AVIATOR (A)	2.0	1.0
4	5	Enter Numbers	AVIATOR (A)	3.0	1.0

Preliminary TAWL Validation Workload Prediction Model - DBH

Segment Decision Rules

Segment 1: Stick (ST)

Continuous Fixed Functions

#	Function Name	Start	Duration	Interrupted By
1	Stick	0.0	-0.5	
6	Monitor Screen	0.0	-180.0	

Preliminary TAWL Validation Workload Prediction Model - DBH

Segment Decision Rules

Segment 2: Dichotic Listening (DL)

Discrete Fixed Functions

#	Function Name	Start	Duration	Interrupted By
4	Detect Numbers	8.0	18.0	
8	Enter Numbers	24.5	4.0	
4	Detect Numbers	28.5	18.0	
8	Enter Numbers	45.0	4.0	
4	Detect Numbers	49.0	18.0	
8	Enter Numbers	65.5	4.0	
4	Detect Numbers	69.5	18.0	
8	Enter Numbers	86.0	4.0	
4	Detect Numbers	90.0	18.0	
8	Enter Numbers	106.5	4.0	
4	Detect Numbers	110.5	18.0	
8	Enter Numbers	127.0	4.0	
4	Detect Numbers	131.0	18.0	
8	Enter Numbers	147.5	4.0	
4	Detect Numbers	151.5	18.0	
8	Enter Numbers	168.0	4.0	

Continuous Fixed Functions

#	Function Name	Start	Duration	Interrupted By
5	Monitor Audio	0.0	-180.0	4

Discrete Random Functions

#	Function Name	Duration	Start	Finish	Times	Interrupted By
7	Enter Number	0.5	11.5	22.5	5	
7	Enter Number	0.5	32.0	43.0	5	
7	Enter Number	0.5	52.5	63.5	5	
7	Enter Number	0.5	73.0	84.0	5	
7	Enter Number	0.5	93.5	104.5	5	
7	Enter Number	0.5	114.0	125.0	5	
7	Enter Number	0.5	134.5	145.5	5	
7	Enter Number	0.5	155.0	166.0	5	

Preliminary TAWL Validation Workload Prediction Model - DBH

Segment Decision Rules

Segment 3: ST and DL

Discrete Fixed Functions

#	Function Name	Start	Duration	Interrupted By
4	Detect Numbers	8.0	18.0	
8	Enter Numbers	24.5	4.0	
4	Detect Numbers	28.5	18.0	
8	Enter Numbers	45.0	4.0	
4	Detect Numbers	49.0	18.0	
8	Enter Numbers	65.5	4.0	
4	Detect Numbers	69.5	18.0	
8	Enter Numbers	86.0	4.0	
4	Detect Numbers	90.0	18.0	
8	Enter Numbers	106.5	4.0	
4	Detect Numbers	110.5	18.0	
8	Enter Numbers	127.0	4.0	
4	Detect Numbers	131.0	18.0	
8	Enter Numbers	147.5	4.0	
4	Detect Numbers	151.5	18.0	
8	Enter Numbers	168.0	4.0	

Continuous Fixed Functions

#	Function Name	Start	Duration	Interrupted By
1	Stick	0.0	-0.5	
6	Monitor Screen	0.0	-180.0	7 8
5	Monitor Audio	0.0	-0.5	4

Discrete Random Functions

#	Function Name	Duration	Start	Finish	Times	Interrupted By
7	Enter Number	0.5	11.5	22.5	5	
7	Enter Number	0.5	32.0	43.0	5	
7	Enter Number	0.5	52.5	63.5	5	
7	Enter Number	0.5	73.0	84.0	5	
7	Enter Number	0.5	93.5	104.5	5	
7	Enter Number	0.5	114.0	125.0	5	
7	Enter Number	0.5	134.5	145.5	5	
7	Enter Number	0.5	155.0	166.0	5	

Preliminary TAWL Validation Workload Prediction Model - DBH

Segment Decision Rules

Segment 4: ST and Pedals (PD)

Continuous Fixed Functions

#	Function Name	Start	Duration	Interrupted By
1	Stick	0.0	-0.5	
2	Pedals	0.0	-0.5	
6	Monitor Screen	0.0	-180.0	

Preliminary TAWL Validation Workload Prediction Model - DBH

Segment Decision Rules

Segment 5: ST, DL, and PD

Discrete Fixed Functions

#	Function Name	Start	Duration	Interrupted By
4	Detect Numbers	8.0	18.0	
8	Enter Numbers	24.5	4.0	
4	Detect Numbers	28.5	18.0	
8	Enter Numbers	45.0	4.0	
4	Detect Numbers	49.0	18.0	
8	Enter Numbers	65.5	4.0	
4	Detect Numbers	69.5	18.0	
8	Enter Numbers	86.0	4.0	
4	Detect Numbers	90.0	18.0	
8	Enter Numbers	106.5	4.0	
4	Detect Numbers	110.5	18.0	
8	Enter Numbers	127.0	4.0	
4	Detect Numbers	131.0	18.0	
8	Enter Numbers	147.5	4.0	
4	Detect Numbers	151.5	18.0	
8	Enter Numbers	168.0	4.0	

Continuous Fixed Functions

#	Function Name	Start	Duration	Interrupted By	
1	Stick	0.0	-0.5		
2	Pedals	0.0	-0.5		
6	Monitor Screen	0.0	-180.0	7	8
5	Monitor Audio	0.0	-0.5	4	

Discrete Random Functions

#	Function Name	Duration	Start	Finish	Times	Interrupted By
7	Enter Number	0.5	11.5	22.5	5	
7	Enter Number	0.5	32.0	43.0	5	
7	Enter Number	0.5	52.5	63.5	5	
7	Enter Number	0.5	73.0	84.0	5	
7	Enter Number	0.5	93.5	104.5	5	
7	Enter Number	0.5	114.0	125.0	5	
7	Enter Number	0.5	134.5	145.5	5	
7	Enter Number	0.5	155.0	166.0	5	

Segment Decision Rules

Segment 6: ST, PD, and Throttle (TR)

Continuous Fixed Functions

#	Function Name	Start	Duration	Interrupted By
1	Stick	0.0	-0.5	
2	Pedals	0.0	-0.5	
3	Throttle	0.0	-0.5	
6	Monitor Screen	0.0	-180.0	

Preliminary TAWL Validation Workload Prediction Model - DBH

Segment Decision Rules

Segment 7: ST, DL, PD, and TR

Discrete Fixed Functions

#	Function Name	Start	Duration	Interrupted By
4	Detect Numbers	8.0	18.0	
8	Enter Numbers	24.5	4.0	
4	Detect Numbers	28.5	18.0	
8	Enter Numbers	45.0	4.0	
4	Detect Numbers	49.0	18.0	
8	Enter Numbers	65.5	4.0	
4	Detect Numbers	69.5	18.0	
8	Enter Numbers	86.0	4.0	
4	Detect Numbers	90.0	18.0	
8	Enter Numbers	106.5	4.0	
4	Detect Numbers	110.5	18.0	
8	Enter Numbers	127.0	4.0	
4	Detect Numbers	131.0	18.0	
8	Enter Numbers	147.5	4.0	
4	Detect Numbers	151.5	18.0	
8	Enter Numbers	168.0	4.0	

Continuous Fixed Functions

#	Function Name	Start	Duration	Interrupted By
1	Stick	0.0	-0.5	
2	Pedals	0.0	-0.5	
3	Throttle	0.0	-0.5	
6	Monitor Screen	0.0	-180.0	7 8
5	Monitor Audio	0.0	-0.5	4

Discrete Random Functions

#	Function Name	Duration	Start	Finish	Times	Interrupted By
7	Enter Number	0.5	11.5	22.5	5	
7	Enter Number	0.5	32.0	43.0	5	
7	Enter Number	0.5	52.5	63.5	5	
7	Enter Number	0.5	73.0	84.0	5	
7	Enter Number	0.5	93.5	104.5	5	
7	Enter Number	0.5	114.0	125.0	5	
7	Enter Number	0.5	134.5	145.5	5	
7	Enter Number	0.5	155.0	166.0	5	

A P P E N D I X   F

CORRELATION MATRIX

Appendix F contains the correlation matrix for 37 variables measured during the TAWL preliminary validation research. The intersection of each row and column of the matrix contains three numbers. The first number is the Pearson product-moment correlation coefficient (*r*) of the row and column variables; the second number is the probability (*p*) of the correlation; the third number is the number (*n*) of pairs used to calculate the correlation.

	Root Mean Squared Error	Omission Error	Commis- sion Error	Exact Match Error	Mental Demand	Physical Demand
Root Mean Squared Error	1.00000 0.0 6	0.99649 0.0533 3	0.78871 0.4215 3	0.99569 0.0592 3	0.97373 0.0010 6	0.97751 0.0008 6
Omission Error	0.99649 0.0533 3	1.00000 0.0 4	0.84551 0.1545 4	0.98140 0.0186 4	0.99815 0.0019 4	0.99338 0.0066 4
Commission Error	0.78871 0.4215 3	0.84551 0.1545 4	1.00000 0.0 4	0.91203 0.0880 4	0.82055 0.1795 4	0.89222 0.1078 4
Exact Match Error	0.99569 0.0592 3	0.98140 0.0186 4	0.91203 0.0880 4	1.00000 0.0 4	0.97767 0.0223 4	0.99687 0.0031 4
Mental Demand	0.97373 0.0010 6	0.99815 0.0019 4	0.82055 0.1795 4	0.97767 0.0223 4	1.00000 0.0 7	0.96984 0.0003 7
Physical Demand	0.97751 0.0008 6	0.99338 0.0066 4	0.89222 0.1078 4	0.99687 0.0031 4	0.96984 0.0003 7	1.00000 0.0 7
Temporal Demand	0.97880 0.0007 6	0.99791 0.0021 4	0.87797 0.1220 4	0.98835 0.0116 4	0.99032 0.0001 7	0.95873 0.0006 7
Own Performance	-0.85684 0.0293 6	-0.95961 0.0404 4	-0.86142 0.1386 4	-0.98659 0.0134 4	-0.92079 0.0032 7	-0.86974 0.0110 7
Effort	0.97128 0.0012 6	0.99122 0.0088 4	0.90705 0.0930 4	0.98942 0.0106 4	0.97031 0.0003 7	0.99360 0.0001 7
Frustration	0.97606 0.0009 6	0.99966 0.0003 4	0.85735 0.1426 4	0.98602 0.0140 4	0.99503 0.0001 7	0.98361 0.0001 7
Task Load Index	0.98965 0.0002 6	0.99913 0.0009 4	0.86658 0.1334 4	0.98785 0.0122 4	0.98377 0.0001 7	0.98850 0.0001 7
Mean Subscale Workload	0.98537 0.0003 6	0.99880 0.0012 4	0.87058 0.1294 4	0.98818 0.0118 4	0.99082 0.0001 7	0.99183 0.0001 7
Time-Based IBI	-0.70355 0.1188 6	-0.74479 0.2552 4	-0.89126 0.1087 4	-0.75446 0.2455 4	-0.68553 0.0891 7	-0.74225 0.0561 7
IBI Standard Deviation	0.08364 0.8748 6	-0.69428 0.3057 4	-0.86405 0.1360 4	-0.81843 0.1816 4	-0.31001 0.4986 7	-0.44021 0.3229 7
IBI Coeff. Variance	0.47264 0.3438 6	-0.52429 0.4757 4	-0.68243 0.3176 4	-0.66650 0.3335 4	-0.02970 0.9496 7	-0.14103 0.7630 7

	Root Mean Squared Error	Omission Error	Commis-sion Error	Exact Match Error	Mental Demand	Physical Demand
RMS Succ. Inter. Differences	0.90783 0.0124 6	0.19788 0.8021 4	-0.16771 0.8323 4	0.01127 0.9887 4	0.49572 0.2579 7	.34723 0.4454 7
RMS Coeff. Variance	0.94123 0.0051 6	0.43161 0.5684 4	0.14741 0.8526 4	0.26863 0.7314 4	0.66583 0.1025 7	0.54566 0.2052 7
Sum Succ. Inter Differences	0.92865 0.0075 6	0.52786 0.4721 4	0.11004 0.8900 4	0.35575 0.6443 4	0.69103 0.0856 7	0.58597 0.1668 7
Decel. / Fluctuations	0.74315 0.0905 6	0.31610 0.6839 4	-0.17469 0.8253 4	0.12961 0.8704 4	0.43370 0.3310 7	0.30798 0.5016 7
Frequency-Based IBI	-0.91684 0.0101 6	-0.96081 0.0392 4	-0.49050 0.5095 4	-0.92826 0.0717 4	-0.86002 0.0130 7	-0.83613 0.0191 7
Respiration Band	-0.72368 0.1040 6	-0.40604 0.5940 4	-0.99551 0.0045 4	-0.22660 0.7734 4	-0.62117 0.1365 7	-0.74963 0.0524 7
Blood Pressure Band	0.30607 0.5552 6	0.04035 0.9596 4	-0.92270 0.0773 4	0.19462 0.8054 4	0.06022 0.8979 7	-0.19370 0.6773 7
Model 1 Cognitive	0.76029 0.0793 6	0.94972 0.0503 4	0.93752 0.0625 4	0.99216 0.0078 4	0.80940 0.0274 7	0.68945 0.0866 7
Model 1 Psychomotor	0.85620 0.0295 6	0.98337 0.0166 4	0.90784 0.0922 4	0.99993 0.0001 4	0.85851 0.0134 7	0.94847 0.0011 7
Model 1 Visual	-0.44989 0.3707 6	0.67257 0.3274 4	0.89232 0.1077 4	0.80206 0.1979 4	0.25576 0.5799 7	0.46989 0.2874 7
Model 1 Auditory	0.44989 0.3707 6	. 4	. 4	. 4	0.30742 0.5024 7	0.10142 0.8287 7
Model 1 Kinesthetic	0.87234 0.0234 6	0.98337 0.0166 4	0.90784 0.0922 4	0.99993 0.0001 4	0.87410 0.0101 7	0.95766 0.0007 7
Model 1 Mean Workload	0.90495 0.0131 6	0.93697 0.0630 4	0.94279 0.0572 4	0.98662 0.0134 4	0.95146 0.0010 7	0.93965 0.0017 7
Model 1 Overload Conditions	0.85581 0.0297 6	0.97689 0.0231 4	0.78305 0.2169 4	0.92365 0.0763 4	0.81812 0.0245 7	0.74946 0.0524 7
Model 1 Overload Density	0.85102 0.0316 6	0.97356 0.0264 4	0.80227 0.1977 4	0.92365 0.0764 4	0.81263 0.0263 7	0.74570 0.0543 7

	Root Mean Squared Error	Omission Error	Commis-sion Error	Exact Match Error	Mental Demand	Physical Demand
Model 2 Cognitive	0.77810 0.0684 6	0.98520 0.0148 4	0.92383 0.0762 4	0.99548 0.0045 4	0.78626 0.0360 7	0.64913 0.1147 7
Model 2 Psychomotor	0.91949 0.0095 6	0.98520 0.0148 4	0.92383 0.0762 4	0.99548 0.0045 4	0.89018 0.0072 7	0.97165 0.0003 7
Model 2 Visual	-0.44989 0.3707 6	0.67257 0.3274 4	0.89232 0.1077 4	0.80206 0.1979 4	0.36306 0.4234 7	0.55187 0.1990 7
Model 2 Auditory	0.44989 0.3707 6	. 4	. 4	. 4	0.30742 0.5024 7	0.10142 0.8287 7
Model 2 Mean Workload	0.88511 0.0190 6	0.93865 0.0614 4	0.96050 0.0395 4	0.98604 0.0140 4	0.93182 0.0022 7	0.90516 0.0051 7
Model 2 Overload Conditions	0.79158 0.0606 6	0.86791 0.1321 4	0.47444 0.5256 4	0.78883 0.2112 4	0.77000 0.0429 7	0.69132 0.0854 7
Model 2 Overload Density	0.79158 0.0606 6	0.86791 0.1321 4	0.47444 0.5256 4	0.78883 0.2112 4	0.77000 0.0429 7	0.69132 0.0854 7

	Temporal Demand	Own Performance	Effort	Frustration	Task Load Index	Mean Subscale Workload
Root Mean Squared Error	0.97880 0.0007 6	-0.85684 0.0293 6	0.97128 0.0012 6	0.97606 0.0009 6	0.98965 0.0002 6	0.98537 0.0003 6
Omission Error	0.99791 0.0021 4	-0.95961 0.0404 4	0.99122 0.0088 4	0.99966 0.0003 4	0.99913 0.0009 4	0.99880 0.0012 4
Commission Error	0.87797 0.1220 4	-0.86142 0.1386 4	0.90705 0.0930 4	0.85735 0.1426 4	0.86658 0.1334 4	0.87058 0.1294 4
Exact Match Error	0.98835 0.0116 4	-0.98659 0.0134 4	0.98942 0.0106 4	0.98602 0.0140 4	0.98785 0.0122 4	0.98818 0.0118 4
Mental Demand	0.99032 0.0001 7	-0.92079 0.0032 7	0.97031 0.0003 7	0.99503 0.0001 7	0.98377 0.0001 7	0.99082 0.0001 7
Physical Demand	0.95873 0.0006 7	-0.86974 0.0110 7	0.99360 0.0001 7	0.98361 0.0001 7	0.98850 0.0001 7	0.99183 0.0001 7
Temporal Demand	1.00000 0.0 7	-0.92248 0.0031 7	0.96527 0.0004 7	0.98732 0.0001 7	0.98217 0.0001 7	0.98487 0.0001 7
Own Performance	-0.92248 0.0031 7	1.00000 0.0 7	-0.89319 0.0068 7	-0.92539 0.0028 7	-0.90591 0.0050 7	-0.89408 0.0066 7
Effort	0.96527 0.0004 7	-0.89319 0.0068 7	1.00000 0.0 7	0.98867 0.0001 7	0.99127 0.0001 7	0.99128 0.0001 7
Frustration	0.98732 0.0001 7	-0.92539 0.0028 7	0.98867 0.0001 7	1.00000 0.0 7	0.99351 0.0001 7	0.99635 0.0001 7
Task Load Index	0.98217 0.0001 7	-0.90591 0.0050 7	0.99127 0.0001 7	0.99351 0.0001 7	1.00000 0.0 7	0.99497 0.0001 7
Mean Subscale Workload	0.98487 0.0001 7	-0.89408 0.0066 7	0.99128 0.0001 7	0.99635 0.0001 7	0.99497 0.0001 7	1.00000 0.0 7
Time-Based IBI	-0.72462 0.0655 7	0.74831 0.0530 7	-0.80040 0.0306 7	-0.74573 0.0543 7	-0.77552 0.0405 7	-0.73372 0.0605 7
IBI Standard Deviation	-0.34022 0.4553 7	0.16367 0.7258 7	-0.36623 0.4191 7	-0.31674 0.4888 7	-0.36737 0.4176 7	-0.37739 0.4040 7
IBI Coeff. Variance	-0.04044 0.9314 7	-0.12354 0.7919 7	-0.04697 0.9204 7	-0.01640 0.9722 7	-0.05312 0.9099 7	-0.07813 0.8678 7

	Temporal Demand	Own Performance	Effort	Frustration	Task Load Index	Mean Subscale Workload
RMS Succ. Inter. Differences	0.46745 0.2902 7	-0.44891 0.3123 7	0.40535 0.3670 7	0.47188 0.2850 7	0.41511 0.3544 7	0.43384 0.3308 7
RMS Coeff. Variance	0.66611 0.1023 7	-0.65826 0.1079 7	0.61409 0.1424 7	0.66066 0.1062 7	0.62474 0.1336 7	0.62372 0.1345 7
Sum Succ. Inter Differences	0.65088 0.1134 7	-0.58623 0.1666 7	0.62373 0.1344 7	0.67441 0.0966 7	0.63944 0.1220 7	0.64961 0.1143 7
Decel. / Fluctuations	0.36243 0.4243 7	-0.26289 0.5690 7	0.31946 0.4849 7	0.39129 0.3854 7	0.33972 0.4560 7	0.37405 0.4085 7
Frequency-Based IBI	-0.92008 0.0033 7	0.85953 0.0132 7	-0.54627 0.2046 7	-0.86802 0.0113 7	-0.88395 0.0083 7	-0.85727 0.0137 7
Respiration Band	-0.71935 0.0684 7	0.51401 0.2379 7	-0.71038 0.0736 7	-0.71333 0.0719 7	-0.73193 0.0615 7	-0.74759 0.0534 7
Blood Pressure Band	-0.08576 0.8549 7	-0.22085 0.6341 7	-0.45385 0.3064 7	-0.08220 0.8609 7	-0.14376 0.7585 7	-0.16529 0.7232 7
Model 1 Cognitive	0.84825 0.0159 7	-0.82580 0.0221 7	0.70068 0.0795 7	0.77370 0.0412 7	0.73831 0.0581 7	0.75661 0.0490 7
Model 1 Psychomotor	0.82679 0.0218 7	-0.77377 0.0412 7	0.94013 0.0016 7	0.89538 0.0064 7	0.90904 0.0046 7	0.90539 0.0050 7
Model 1 Visual	0.25075 0.5876 7	-0.14879 0.7502 7	0.41710 0.3519 7	0.30939 0.4995 7	0.36696 0.4181 7	0.36184 0.4251 7
Model 1 Auditory	0.36601 0.4194 7	-0.38814 0.3896 7	0.12552 0.7886 7	0.24189 0.6013 7	0.18924 0.6845 7	0.21036 0.6499 7
Model 1 Kinesthetic	0.84416 0.0169 7	-0.79156 0.0339 7	0.95008 0.0010 7	0.90898 0.0046 7	0.92096 0.0032 7	0.91801 0.0035 7
Model 1 Mean Workload	0.96578 0.0004 7	-0.91112 0.0043 7	0.93493 0.0020 7	0.95005 0.0010 7	0.94058 0.0016 7	0.95120 0.0010 7
Model 1 Overload Conditions	0.84835 0.0158 7	-0.61583 0.1409 7	0.73691 0.0588 7	0.78230 0.0376 7	0.78602 0.0361 7	0.80514 0.0289 7
Model 1 Overload Density	0.84726 0.0161 7	-0.61354 0.1428 7	0.73607 0.0593 7	0.77870 0.0391 7	0.78233 0.0376 7	0.80190 0.0301 7

	Temporal Demand	Own Performance	Effort	Frustration	Task Load Index	Mean Subscale Workload
Model 2 Cognitive	0.82731 0.0216 7	-0.79544 0.0324 7	0.66699 0.1017 7	0.74671 0.0538 7	0.70843 0.0748 7	0.72753 0.0639 7
Model 2 Psychomotor	0.87056 0.0108 7	-0.78074 0.0383 7	0.96526 0.0004 7	0.92317 0.0030 7	0.93839 0.0018 7	0.93800 0.0018 7
Model 2 Visual	0.37201 0.4112 7	-0.26323 0.5684 7	0.49870 0.2546 7	0.40689 0.3650 7	0.45833 0.3010 7	0.45797 0.3014 7
Model 2 Auditory	0.36601 0.4194 7	-0.38814 0.3896 7	0.12552 0.7886 7	0.24189 0.6013 7	0.18924 0.6845 7	0.21086 0.6499 7
Model 2 Mean Workload	0.95870 0.0007 7	-0.88023 0.0089 7	0.90220 0.0054 7	0.92403 0.0029 7	0.91408 0.0040 7	0.92789 0.0026 7
Model 2 Overload Conditions	0.74752 0.0534 7	-0.55917 0.1919 7	0.64811 0.1154 7	0.71802 0.0692 7	0.72219 0.0668 7	0.73379 0.0605 7
Model 2 Overload Density	0.74752 0.0534 7	-0.55917 0.1919 7	0.64811 0.1154 7	0.71802 0.0692 7	0.72219 0.0668 7	0.73379 0.0605 7

	Time-Based IBI	IBI Standard Deviation	IBI Coeff. Variance	RMS Succ. Inter. Diff.	RMS Coeff. Variance
Root Mean Squared Error	-0.70355 0.1188 6	0.08364 0.8748 6	0.47264 0.3438 6	0.90783 0.0124 6	0.94123 0.0051 6
Omission Error	-0.74479 0.2552 4	-0.69428 0.3057 4	-0.52429 0.4757 4	0.19788 0.8021 4	0.43161 0.5684 4
Commission Error	-0.89126 0.1087 4	-0.86405 0.1360 4	-0.68243 0.3176 4	-0.16771 0.8323 4	0.14741 0.8526 4
Exact Match Error	-0.75446 0.2455 4	-0.81643 0.1816 4	-0.66650 0.3335 4	0.01127 0.9887 4	0.26863 0.7314 4
Mental Demand	-0.68553 0.0891 7	-0.31001 0.4986 7	-0.02970 0.9496 7	0.49572 0.2579 7	0.66583 0.1025 7
Physical Demand	-0.74225 0.0561 7	-0.44021 0.3229 7	-0.14103 0.7630 7	0.34723 0.4454 7	0.54566 0.2052 7
Temporal Demand	-0.72462 0.0655 7	-0.34022 0.4553 7	-0.04044 0.9314 7	0.46745 0.2902 7	0.66611 0.1023 7
Own Performance	0.74831 0.0530 7	0.16367 0.7258 7	-0.12354 0.7919 7	-0.44891 0.3123 7	-0.65826 0.1079 7
Effort	-0.80040 0.0306 7	-0.36623 0.4191 7	-0.04697 0.9204 7	0.40535 0.3670 7	0.61409 0.1424 7
Frustration	-0.74573 0.0543 7	-0.31674 0.4888 7	-0.01640 0.9722 7	0.47188 0.2850 7	0.66066 0.1062 7
Task Load Index	-0.77552 0.0405 7	-0.36737 0.4176 7	-0.05312 0.9099 7	0.41511 0.3544 7	0.62474 0.1336 7
Mean Subscale Workload	-0.73372 0.0605 7	-0.37739 0.4040 7	-0.07813 0.8678 7	0.43384 0.3308 7	0.62372 0.1345 7
Time-Based IBI	1.00000 0.0 7	0.11577 0.8048 7	-0.26558 0.5649 7	-0.32265 0.4803 7	-0.60974 0.1460 7
IBI Standard Deviation	0.11577 0.8048 7	1.00000 0.0 7	0.92505 0.0028 7	0.60296 0.1518 7	0.44055 0.3225 7
IBI Coeff. Variance	-0.26558 0.5649 7	0.92505 0.0028 7	1.00000 0.0 7	0.73011 0.0625 7	0.68084 0.0922 7

	Time-Based IBI	IBI Standard Deviation	IBI Coeff. Variance	RMS Succ. Inter. Diff.	RMS Coeff. Variance
RMS Succ. Inter. Differences	-0.32265 0.4803 7	0.60296 0.1518 7	0.73011 0.0625 7	1.00000 0.0 7	0.93931 0.0017 7
RMS Coeff. Variance	-0.60974 0.1460 7	0.44055 0.3225 7	0.68084 0.0922 7	0.93931 0.0017 7	1.00000 0.0 7
Sum Succ. Inter Differences	-0.46386 0.2944 7	0.38142 0.3985 7	0.56767 0.1837 7	0.94113 0.0016 7	0.93271 0.0022 7
Decel. / Fluctuations	-0.03918 0.9335 7	0.46850 0.2890 7	0.49244 0.2616 7	0.88712 0.0077 7	0.73096 0.0620 7
Frequency-Based IBI	0.60195 0.1527 7	0.12708 0.7860 7	-0.05658 0.9041 7	0.09270 0.8433 7	-0.00582 0.9901 7
Respiration Band	0.06819 0.8845 7	0.77798 0.0394 7	0.66891 0.1004 7	0.85816 0.0135 7	0.81672 0.0250 7
Blood Pressure Band	-0.34919 0.4427 7	0.95751 0.0007 7	0.93509 0.0020 7	0.93955 0.0017 7	0.91856 0.0035 7
Model 1 Cognitive	-0.44167 0.3211 7	-0.25188 0.5858 7	-0.05576 0.9055 7	0.42373 0.3435 7	0.57319 0.1786 7
Model 1 Psychomotor	-0.75610 0.0492 7	-0.42798 0.3381 7	-0.13871 0.7668 7	0.22308 0.6306 7	0.41671 0.3524 7
Model 1 Visual	-0.32958 0.4704 7	-0.82675 0.0218 7	-0.69878 0.0807 7	-0.57909 0.1731 7	-0.40435 0.3683 7
Model 1 Auditory	0.02333 0.9604 7	0.10661 0.8200 7	0.12529 0.7890 7	0.39521 0.3802 7	0.41312 0.3569 7
Model 1 Kinesthetic	-0.75992 0.0474 7	-0.42699 0.3393 7	-0.13533 0.7724 7	0.23776 0.6077 7	0.43317 0.3316 7
Model 1 Mean Workload	-0.66460 0.1034 7	-0.49191 0.2622 7	-0.21809 0.6385 7	0.28299 0.5386 7	0.49506 0.2586 7
Model 1 Overload Conditions	-0.41358 0.3564 7	-0.37967 0.4009 7	-0.17055 0.7147 7	0.45840 0.3009 7	0.57406 0.1777 7
Model 1 Overload Density	-0.42728 0.3390 7	-0.38088 0.3993 7	-0.16624 0.7217 7	0.45611 0.3037 7	0.57740 0.1746 7

	Time-Based IBI	IBI Standard Deviation	IBI Coeff. Variance	RMS Succ. Inter. Diff.	RMS Coeff. Variance
Model 2 Cognitive	-0.42362 0.3436 7	-0.16732 0.7199 7	0.02386 0.9595 7	0.50079 0.2523 7	0.63273 0.1272 7
Model 2 Psychomotor	-0.77316 0.0415 7	-0.45667 0.3030 7	-0.15313 0.7431 7	0.25719 0.5777 7	0.45890 0.3003 7
Model 2 Visual	-0.36361 0.4227 7	-0.90015 0.0057 7	-0.75208 0.0512 7	-0.55085 0.2000 7	-0.35053 0.4408 7
Model 2 Auditory	0.02333 0.9604 7	0.10661 0.8200 7	0.12529 0.7890 7	0.39521 0.3802 7	0.41312 0.3569 7
Model 2 Mean Workload	-0.63179 0.1280 7	-0.49574 0.2579 7	-0.22801 0.6229 7	0.29469 0.5212 7	0.50348 0.2493 7
Model 2 Overload Conditions	-0.20688 0.6563 7	-0.31574 0.4903 7	-0.19553 0.6744 7	0.42271 0.3447 7	0.46074 0.2981 7
Model 2 Overload Density	-0.20688 0.6563 7	-0.31574 0.4903 7	-0.19553 0.6744 7	0.42271 0.3447 7	0.46074 0.2981 7

	Sum Succ. Inter Diff.	Decel. / Fluct.	Frequency-Based IBI	Respir-ation Band	Blood Pressure Band
Root Mean Squared Error	0.92865 0.0075 6	0.74315 0.0905 6	-0.91684 0.0101 6	-0.72368 0.1040 6	0.30607 0.5552 6
Omission Error	0.52786 0.4721 4	0.31610 0.6839 4	-0.96081 0.0392 4	-0.40604 0.5940 4	0.04035 0.9596 4
Commission Error	0.11004 0.8900 4	-0.17469 0.8253 4	-0.49050 0.5095 4	-0.99551 0.0045 4	-0.92270 0.0773 4
Exact Match Error	0.35575 0.6443 4	0.12961 0.8704 4	-0.92826 0.0717 4	-0.22660 0.7734 4	0.19462 0.8054 4
Mental Demand	0.69103 0.0856 7	0.43370 0.3310 7	-0.86002 0.0130 7	-0.62117 0.1365 7	0.06022 0.8979 7
Physical Demand	0.58597 0.1668 7	0.30798 0.5016 7	-0.83613 0.0191 7	-0.74963 0.0524 7	-0.19370 0.6773 7
Temporal Demand	0.65088 0.1134 7	0.36243 0.4243 7	-0.92008 0.0033 7	-0.71935 0.0684 7	-0.08576 0.8549 7
Own Performance	-0.58623 0.1666 7	-0.26289 0.5690 7	0.85953 0.0132 7	0.51401 0.2379 7	-0.22085 0.6341 7
Effort	0.62373 0.1344 7	0.31946 0.4849 7	-0.54627 0.2046 7	-0.71038 0.0736 7	-0.45385 0.3064 7
Frustration	0.67441 0.0966 7	0.39129 0.3854 7	-0.86802 0.0113 7	-0.71333 0.0719 7	-0.08220 0.8609 7
Task Load Index	0.63944 0.1220 7	0.33972 0.4560 7	-0.88395 0.0083 7	-0.73193 0.0615 7	-0.14376 0.7585 7
Mean Subscale Workload	0.64961 0.1143 7	0.37405 0.4085 7	-0.85727 0.0137 7	-0.74759 0.0534 7	-0.16529 0.7232 7
Time-Based IBI	-0.46386 0.2944 7	-0.03918 0.9335 7	0.60195 0.1527 7	0.06819 0.8845 7	-0.34919 0.4427 7
IBI Standard Deviation	0.38142 0.3985 7	0.46850 0.2890 7	0.12708 0.7860 7	0.77798 0.0394 7	0.95751 0.0007 7
IBI Coeff. Variance	0.56767 0.1837 7	0.49244 0.2616 7	-0.05658 0.9041 7	0.66891 0.1004 7	0.93509 0.0020 7

	Sum Succ. Inter Diff.	Decel. / Fluct.	Frequency-Based IBI	Respir-ation Band	Blood Pressure Band
RMS Succ. Inter. Differences	0.94113 0.0016 7	0.88712 0.0077 7	0.09270 0.8433 7	0.85816 0.0135 7	0.93955 0.0017 7
RMS Coeff. Variance	0.93271 0.0022 7	0.73096 0.0620 7	-0.00582 0.9901 7	0.81672 0.0250 7	0.91856 0.0035 7
Sum Succ. Inter Differences	1.00000 0.0 7	0.89839 0.0060 7	-0.15836 0.7345 7	0.62759 0.1313 7	0.81610 0.0252 7
Decel. / Fluctuations	0.89839 0.0060 7	1.00000 0.0 7	-0.08987 0.8480 7	0.62742 0.1315 7	0.78590 0.0362 7
Frequency-Based IBI	-0.15836 0.7345 7	-0.08987 0.8480 7	1.00000 0.0 7	0.55278 0.1981 7	-0.08629 0.8540 7
Respiration Band	0.62759 0.1313 7	0.62742 0.1315 7	0.55278 0.1981 7	1.00000 0.0 7	0.69432 0.0835 7
Blood Pressure Band	0.81610 0.0252 7	0.78590 0.0362 7	-0.08629 0.8540 7	0.69432 0.0835 7	1.00000 0.0 7
Model 1 Cognitive	0.46624 0.2916 7	0.24896 0.5903 7	-0.89215 0.0069 7	-0.77365 0.0413 7	-0.18902 0.6848 7
Model 1 Psychomotor	0.48757 0.2670 7	0.21589 0.6420 7	-0.67038 0.0993 7	-0.73944 0.0575 7	-0.17293 0.7108 7
Model 1 Visual	-0.32277 0.4801 7	-0.46360 0.2947 7	-0.74066 0.0569 7	-0.66817 0.1009 7	-0.21364 0.6455 7
Model 1 Auditory	0.25102 0.5871 7	0.20258 0.6631 7	-0.07799 0.8680 7	-0.58591 0.1669 7	-0.73472 0.0600 7
Model 1 Kinesthetic	0.49900 0.2543 7	0.22403 0.6291 7	-0.25779 0.5768 7	-0.29410 0.5220 7	0.05270 0.9107 7
Model 1 Mean Workload	0.46137 0.2974 7	0.17709 0.7041 7	-0.77560 0.0404 7	-0.88717 0.0077 7	-0.33652 0.4605 7
Model 1 Overload Conditions	0.60328 0.1515 7	0.45370 0.3065 7	-0.75386 0.0503 7	-0.68215 0.0914 7	-0.21317 0.6463 7
Model 1 Overload Density	0.59403 0.1596 7	0.43645 0.3276 7	-0.79671 0.0320 7	-0.72213 0.0669 7	-0.29020 0.5278 7

	Sum Succ. Inter Diff.	Decel. / Fluct.	Frequency-Based IBI	Respir-ation Band	Blood Pressure Band
Model 2 Cognitive	0.52100 0.2305 7	0.31679 0.4888 7	-0.67473 0.0964 7	-0.67235 0.0980 7	-0.10537 0.8221 7
Model 2 Psychomotor	0.51635 0.2354 7	0.23638 0.6098 7	-0.77325 0.0414 7	-0.71271 0.0723 7	-0.26517 0.5655 7
Model 2 Visual	-0.29939 0.5142 7	-0.46927 0.2881 7	-0.15148 0.7458 7	-0.72979 0.0626 7	-0.81040 0.0271 7
Model 2 Auditory	0.25102 0.5871 7	0.20258 0.6631 7	-0.25779 0.5768 7	-0.29410 0.5220 7	0.05270 0.9107 7
Model 2 Mean Workload	0.45169 0.3090 7	0.17535 0.7069 7	-0.75949 0.0476 7	-0.89506 0.0065 7	-0.37091 0.4127 7
Model 2 Overload Conditions	0.62528 0.1332 7	0.58416 0.1685 7	-0.82035 0.0238 7	-0.61407 0.1424 7	-0.09440 0.8405 7
Model 2 Overload Density	0.62528 0.1332 7	0.58416 0.1685 7	-0.82035 0.0238 7	-0.61407 0.1424 7	-0.09440 0.8405 7

	Model 1 Cognitive	Model 1 Psychomotor	Model 1 Visual	Model 1 Auditory
Root Mean Squared Error	0.76029 0.0793 6	0.85620 0.0295 6	-0.44989 0.3707 6	0.44989 0.3707 6
Omission Error	0.94972 0.0503 4	0.98337 0.0166 4	0.67257 0.3274 4	. 4
Commission Error	0.93752 0.0625 4	0.90784 0.0922 4	0.89232 0.1077 4	. 4
Exact Match Error	0.99216 0.0078 4	0.99993 0.0001 4	0.80206 0.1979 4	. 4
Mental Demand	0.80940 0.0274 7	0.85851 0.0134 7	0.25576 0.5799 7	0.30742 0.5024 7
Physical Demand	0.68945 0.0866 7	0.94847 0.0011 7	0.46989 0.2874 7	0.10142 0.8287 7
Temporal Demand	0.84825 0.0159 7	0.82679 0.0218 7	0.25075 0.5876 7	0.36601 0.4194 7
Own Performance	-0.82580 0.0221 7	-0.77377 0.0412 7	-0.14879 0.7502 7	-0.38814 0.3896 7
Effort	0.70068 0.0795 7	0.94013 0.0016 7	0.41710 0.3519 7	0.12552 0.7886 7
Frustration	0.77370 0.0412 7	0.89538 0.0064 7	0.30939 0.4995 7	0.24189 0.6013 7
Task Load Index	0.73831 0.0581 7	0.90904 0.0046 7	0.36696 0.4181 7	0.18924 0.6845 7
Mean Subscale Workload	0.75661 0.0490 7	0.90539 0.0050 7	0.36184 0.4251 7	0.21086 0.6499 7
Time-Based IBI	-0.44167 0.3211 7	-0.75610 0.0492 7	-0.32958 0.4704 7	0.02333 0.9604 7
IBI Standard Deviation	-0.25188 0.5858 7	-0.42798 0.3381 7	-0.82675 0.0218 7	0.10661 0.8200 7
IBI Coeff. Variance	-0.05576 0.9055 7	-0.13871 0.7668 7	-0.69878 0.0807 7	0.12529 0.7890 7

	Model 1 Cognitive	Model 1 Psychomotor	Model 1 Visual	Model 1 Auditory
RMS Succ. Inter. Differences	0.42373 0.3435 7	0.22308 0.6306 7	-0.57909 0.1731 7	0.39521 0.3802 7
RMS Coeff. Variance	0.57319 0.1786 7	0.41671 0.3524 7	-0.40435 0.3683 7	0.41312 0.3569 7
Sum Succ. Inter Differences	0.46624 0.2916 7	0.48757 0.2670 7	-0.32277 0.4801 7	0.25102 0.5871 7
Decel. / Fluctuations	0.24896 0.5903 7	0.21589 0.6420 7	-0.46360 0.2947 7	0.20258 0.6631 7
Frequency-Based IBI	-0.89215 0.0069 7	-0.67038 0.0993 7	-0.74066 0.0569 7	-0.07799 0.8680 7
Respiration Band	-0.77365 0.0413 7	-0.73944 0.0575 7	-0.66817 0.1009 7	-0.58591 0.1669 7
Blood Pressure Band	-0.18902 0.6848 7	-0.17293 0.7108 7	-0.21364 0.6455 7	-0.73472 0.0600 7
Model 1 Cognitive	1.00000 0.0 7	0.44994 0.3111 7	-0.06653 0.8873 7	0.78239 0.0376 7
Model 1 Psychomotor	0.44994 0.3111 7	1.00000 0.0 7	0.62483 0.1336 7	-0.19682 0.6723 7
Model 1 Visual	-0.06653 0.8873 7	0.62483 0.1336 7	1.00000 0.0 7	-0.57085 0.1807 7
Model 1 Auditory	0.78239 0.0376 7	-0.19682 0.6723 7	-0.57085 0.1807 7	1.00000 0.0 7
Model 1 Kinesthetic	0.47905 0.2767 7	0.99945 0.0001 7	0.60939 0.1463 7	-0.16432 0.7248 7
Model 1 Mean Workload	0.87052 0.0108 7	0.81772 0.0246 7	0.37369 0.4090 7	0.37571 0.4062 7
Model 1 Overload Conditions	0.79786 0.0315 7	0.52458 0.2267 7	0.07830 0.8675 7	0.51096 0.2412 7
Model 1 Overload Density	0.80294 0.0297 7	0.51920 0.2324 7	0.07954 0.8654 7	0.51904 0.2326 7

	Model 1 Cognitive	Model 1 Psychomotor	Model 1 Visual	Model 1 Auditory
Model 2 Cognitive	0.99330 0.0001 7	0.39409 0.3817 7	-0.16231 0.7281 7	0.82138 0.0235 7
Model 2 Psychomotor	0.51248 0.2396 7	0.99131 0.0001 7	0.60447 0.1505 7	-0.12464 0.7900 7
Model 2 Visual	0.11987 0.7980 7	0.65141 0.1130 7	0.97896 0.0001 7	-0.39130 0.3854 7
Model 2 Auditory	0.78239 0.0376 7	-0.19682 0.6723 7	-0.57085 0.1807 7	1.00000 0.0001 7
Model 2 Mean Workload	0.90973 0.0045 7	0.75014 0.0521 7	0.31739 0.4879 7	0.46261 0.2959 7
Model 2 Overload Conditions	0.63542 0.1251 7	0.51426 0.2377 7	0.05418 0.9082 7	0.35355 0.4366 7
Model 2 Overload Density	0.63542 0.1251 7	0.51426 0.2377 7	0.05418 0.9082 7	0.35355 0.4366 7

	Model 1 Kinesthetic	Model 1 Mean Workload	Model 1 Overload Condtions	Model 1 Overload Density
Root Mean Squared Error	0.87234 0.0234 6	0.90495 0.0131 6	0.85581 0.0297 6	0.85102 0.0316 6
Omission Error	0.98337 0.0166 4	0.93697 0.0630 4	0.97689 0.0231 4	0.97356 0.0264 4
Commission Error	0.90784 0.0922 4	0.94279 0.0572 4	0.78305 0.2169 4	0.80227 0.1977 4
Exact Match Error	0.99993 0.0001 4	0.98662 0.0134 4	0.92365 0.0763 4	0.92365 0.0764 4
Mental Demand	0.87410 0.0101 7	0.95146 0.0010 7	0.81812 0.0245 7	0.81263 0.0263 7
Physical Demand	0.95766 0.0007 7	0.93965 0.0017 7	0.74946 0.0524 7	0.74570 0.0543 7
Temporal Demand	0.84416 0.0169 7	0.96578 0.0004 7	0.84835 0.0158 7	0.84726 0.0161 7
Own Performance	-0.79156 0.0339 7	-0.91112 0.0043 7	-0.61583 0.1409 7	-0.61354 0.1428 7
Effort	0.95008 0.0010 7	0.93493 0.0020 7	0.73691 0.0588 7	0.73607 0.0593 7
Frustration	0.90898 0.0046 7	0.95005 0.0010 7	0.78230 0.0376 7	0.77870 0.0391 7
Task Load Index	0.92096 0.0032 7	0.94058 0.0016 7	0.78602 0.0361 7	0.78233 0.0376 7
Mean Subscale Workload	0.91801 0.0035 7	0.95120 0.0010 7	0.80514 0.0289 7	0.80190 0.0301 7
Time-Based IBI	-0.75992 0.0474 7	-0.66460 0.1034 7	-0.41358 0.3564 7	-0.42728 0.3390 7
IBI Standard Deviation	-0.42699 0.3393 7	-0.49191 0.2622 7	-0.37967 0.4009 7	-0.38088 0.3993 7
IBI Coeff. Variance	-0.13533 0.7724 7	-0.21809 0.6385 7	-0.17055 0.7147 7	-0.16624 0.7217 7

	Model 1 Kinesthetic	Model 1 Mean Workload	Model 1 Overload Condtions	Model 1 Overload Density
RMS Succ. Inter. Differences	0.23776 0.6077 7	0.28299 0.5386 7	0.45840 0.3009 7	0.45611 0.3037 7
RMS Coeff. Variance	0.43317 0.3316 7	0.49506 0.2586 7	0.57406 0.1777 7	0.57740 0.1746 7
Sum Succ. Inter Differences	0.49900 0.2543 7	0.46137 0.2974 7	0.60328 0.1515 7	0.59403 0.1596 7
Decel. / Fluctuations	0.22403 0.6291 7	0.17709 0.7041 7	0.45370 0.3065 7	0.43645 0.3276 7
Frequency-Based IBI	-0.25779 0.5768 7	-0.77560 0.0404 7	-0.75386 0.0503 7	-0.79671 0.0320 7
Respiration Band	-0.29410 0.5220 7	-0.88717 0.0077 7	-0.68215 0.0914 7	-0.72213 0.0669 7
Blood Pressure Band	0.05270 0.9107 7	-0.33652 0.4605 7	-0.21317 0.6463 7	-0.29020 0.5278 7
Model 1 Cognitive	0.47905 0.2767 7	0.87052 0.0108 7	0.79786 0.0315 7	0.80294 0.0297 7
Model 1 Psychomotor	0.99945 0.0001 7	0.81772 0.0246 7	0.52458 0.2267 7	0.51920 0.2324 7
Model 1 Visual	0.60939 0.1463 7	0.37369 0.4090 7	0.07830 0.8675 7	0.07954 0.8654 7
Model 1 Auditory	-0.16432 0.7248 7	0.37571 0.4062 7	0.51096 0.2412 7	0.51904 0.2326 7
Model 1 Kinesthetic	1.00000 0.0 7	0.83536 0.0193 7	0.54500 0.2058 7	0.53986 0.2110 7
Model 1 Mean Workload	0.83536 0.0193 7	1.00000 0.0 7	0.78219 0.0377 7	0.78392 0.0370 7
Model 1 Overload Conditions	0.54500 0.2058 7	0.78219 0.0377 7	1.00000 0.0 7	0.99906 0.0001 7
Model 1 Overload Density	0.53986 0.2110 7	0.78392 0.0370 7	0.99906 0.0001 7	1.00000 0.0 7

	Model 1 Kinesthetic	Model 1 Mean Workload	Model 1 Overload Condtions	Model 1 Overload Density
Model 2 Cognitive	0.42418 0.3429 7	0.82702 0.0217 7	0.81113 0.0268 7	0.81686 0.0249 7
Model 2 Psychomotor	0.99314 0.0001 7	0.85499 0.0142 7	0.61723 0.1398 7	0.61466 0.1419 7
Model 2 Visual	0.64218 0.1199 7	0.51222 0.2399 7	0.21475 0.6438 7	0.21815 0.6384 7
Model 2 Auditory	-0.16432 0.7248 7	0.37571 0.4062 7	0.51096 0.2412 7	0.51904 0.2326 7
Model 2 Mean Workload	0.77030 0.0427 7	0.99152 0.0001 7	0.83156 0.0204 7	0.83594 0.0191 7
Model 2 Overload Conditions	0.52931 0.2218 7	0.65886 0.1075 7	0.87727 0.0095 7	0.85562 0.0141 7
Model 2 Overload Density	0.52931 0.2218 7	0.65886 0.1075 7	0.87727 0.0095 7	0.85562 0.0141 7

	Model 2 Cognitive	Model 2 Psychomotor	Model 2 Visual	Model 2 Auditory
Root Mean Squared Error	0.77810 0.0684 6	0.91949 0.0095 6	-0.44989 0.3707 6	0.44989 0.3707 6
Omission Error	0.98520 0.0148 4	0.98520 0.0148 4	0.67257 0.3274 4	. 4
Commission Error	0.92383 0.0762 4	0.92383 0.0762 4	0.89232 0.1077 4	. 4
Exact Match Error	0.99548 0.0045 4	0.99548 0.0045 4	0.80206 0.1979 4	. 4
Mental Demand	0.78626 0.0360 7	0.89018 0.0072 7	0.36306 0.4234 7	0.30742 0.5024 7
Physical Demand	0.64913 0.1147 7	0.97165 0.0003 7	0.55187 0.1990 7	0.10142 0.8287 7
Temporal Demand	0.82731 0.0216 7	0.87056 0.0108 7	0.37201 0.4112 7	0.36601 0.4194 7
Own Performance	-0.79544 0.0324 7	-0.78074 0.0383 7	-0.26323 0.5684 7	-0.38814 0.3896 7
Effort	0.66699 0.1017 7	0.96526 0.0004 7	0.49870 0.2546 7	0.12552 0.7886 7
Frustration	0.74671 0.0538 7	0.92317 0.0030 7	0.40689 0.3650 7	0.24189 0.6013 7
Task Load Index	0.70843 0.0748 7	0.93839 0.0018 7	0.45833 0.3010 7	0.18924 0.6845 7
Mean Subscale Workload	0.72753 0.0639 7	0.93800 0.0018 7	0.45797 0.3014 7	0.21086 0.6499 7
Time-Based IBI	-0.42362 0.3436 7	-0.77316 0.0415 7	-0.36361 0.4227 7	0.02333 0.9604 7
IBI Standard Deviation	-0.16732 0.7199 7	-0.45667 0.3030 7	-0.90015 0.0057 7	0.10661 0.8200 7
IBI Coeff. Variance	0.02386 0.9595 7	-0.15313 0.7431 7	-0.75208 0.0512 7	0.12529 0.7890 7

	Model 2 Cognitive	Model 2 Psychomotor	Model 2 Visual	Model 2 Auditory
RMS Succ. Inter. Differences	0.50079 0.2523 7	0.25719 0.5777 7	-0.55085 0.2000 7	0.39521 0.3802 7
RMS Coeff. Variance	0.63273 0.1272 7	0.45890 0.3003 7	-0.35053 0.4408 7	0.41312 0.3569 7
Sum Succ. Inter Differences	0.52100 0.2305 7	0.51635 0.2354 7	-0.29939 0.5142 7	0.25102 0.5871 7
Decel. / Fluctuations	0.31679 0.4888 7	0.23638 0.6098 7	-0.46927 0.2881 7	0.20258 0.6631 7
Frequency-Based IBI	-0.67473 0.0964 7	-0.77325 0.0414 7	-0.15148 0.7458 7	-0.25779 0.5768 7
Respiration Band	-0.67235 0.0980 7	-0.71271 0.0723 7	-0.72979 0.0626 7	-0.29410 0.5220 7
Blood Pressure Band	-0.10537 0.8221 7	-0.26517 0.5655 7	-0.81040 0.0271 7	0.05270 0.9107 7
Model 1 Cognitive	0.99330 0.0001 7	0.51248 0.2396 7	0.11987 0.7980 7	0.78239 0.0376 7
Model 1 Psychomotor	0.39409 0.3817 7	0.99131 0.0001 7	0.65141 0.1130 7	-0.19682 0.6723 7
Model 1 Visual	-0.16231 0.7281 7	0.60447 0.1505 7	0.97896 0.0001 7	-0.57085 0.1807 7
Model 1 Auditory	0.82138 0.0235 7	-0.12464 0.7900 7	-0.39130 0.3854 7	1.00000 0.0001 7
Model 1 Kinesthetic	0.42418 0.3429 7	0.99314 0.0001 7	0.64218 0.1199 7	-0.16432 0.7248 7
Model 1 Mean Workload	0.82702 0.0217 7	0.85499 0.0142 7	0.51222 0.2399 7	0.37571 0.4062 7
Model 1 Overload Conditions	0.81113 0.0268 7	0.61723 0.1398 7	0.21475 0.6438 7	0.51096 0.2412 7
Model 1 Overload Density	0.81686 0.0249	0.61466 0.1419	0.21815 0.6384	0.51904 0.2326

	Model 2 Cognitive	Model 2 Psychomotor	Model 2 Visual	Model 2 Auditory
Model 2 Cognitive	7 1.00000 0.0 7	7 0.46355 0.2948 7	7 0.02221 0.9623 7	7 0.82138 0.0235 7
Model 2 Psychomotor	0.46355 0.2948 7	1.00000 0.0 7	0.64653 0.1166 7	-0.12464 0.7900 7
Model 2 Visual	0.02221 0.9623 7	0.64653 0.1166 7	1.00000 0.0 7	-0.39130 0.3854 7
Model 2 Auditory	0.82138 0.0235 7	-0.12464 0.7900 7	-0.39130 0.3854 7	1.00000 0.0 7
Model 2 Mean Workload	0.87438 0.0100 7	0.80236 0.0299 7	0.47071 0.2864 7	0.46261 0.2959 7
Model 2 Overload Conditions	0.63969 0.1218 7	0.56353 0.1877 7	0.14859 0.7505 7	0.35355 0.4366 7
Model 2 Overload Density	0.63969 0.1218 7	0.56353 0.1877 7	0.14859 0.7505 7	0.35355 0.4366 7

	Model 2 Mean Workload	Model 2 Overload Conditions	Model 2 Overload Density
Root Mean Squared Error	0.88511 0.0190 6	0.79158 0.0606 6	0.79158 0.0606 6
Omission Error	0.93865 0.0614 4	0.86791 0.1321 4	0.86791 0.1321 4
Commission Error	0.96050 0.0395 4	0.47444 0.5256 4	0.47444 0.5256 4
Exact Match Error	0.98604 0.0140 4	0.78883 0.2112 4	0.78883 0.2112 4
Mental Demand	0.93182 0.0022 7	0.77000 0.0429 7	0.77000 0.0429 7
Physical Demand	0.90516 0.0051 7	0.69132 0.0854 7	0.69132 0.0854 7
Temporal Demand	0.95870 0.0007 7	0.74752 0.0534 7	0.74752 0.0534 7
Own Performance	-0.88023 0.0089 7	-0.55917 0.1919 7	-0.55917 0.1919 7
Effort	0.90220 0.0054 7	0.64811 0.1154 7	0.64811 0.1154 7
Frustration	0.92403 0.0029 7	0.71802 0.0692 7	0.71802 0.0692 7
Task Load Index	0.91408 0.0040 7	0.72219 0.0668 7	0.72219 0.0668 7
Mean Subscale Workload	0.92789 0.0026 7	0.73379 0.0605 7	0.73379 0.0605 7
Time-Based IBI	-0.63179 0.1280 7	-0.20688 0.6563 7	-0.20688 0.6563 7
IBI Standard Deviation	-0.49574 0.2579 7	-0.31574 0.4903 7	-0.31574 0.4903 7
IBI Coeff. Variance	-0.22801 0.6229 7	-0.19553 0.6744 7	-0.19553 0.6744 7

	Model 2 Mean Workload	Model 2 Overload Conditions	Model 2 Overload Density
RMS Succ. Inter. Differences	0.29469 0.5212 7	0.42271 0.3447 7	0.42271 0.3447 7
RMS Coeff. Variance	0.50348 0.2493 7	0.46074 0.2981 7	0.46074 0.2981 7
Sum Succ. Inter Differences	0.45169 0.3090 7	0.62528 0.1332 7	0.62528 0.1332 7
Decel. / Fluctuations	0.17535 0.7069 7	0.58416 0.1685 7	0.58416 0.1685 7
Frequency-Based IBI	-0.75949 0.0476 7	-0.82035 0.0238 7	-0.82035 0.0238 7
Respiration Band	-0.89506 0.0065 7	-0.61407 0.1424 7	-0.61407 0.1424 7
Blood Pressure Band	-0.37091 0.4127 7	-0.09440 0.8405 7	-0.09440 0.8405 7
Model 1 Cognitive	0.90973 0.0045 7	0.63542 0.1251 7	0.63542 0.1251 7
Model 1 Psychomotor	0.75014 0.0521 7	0.51426 0.2377 7	0.51426 0.2377 7
Model 1 Visual	0.31739 0.4879 7	0.05418 0.9082 7	0.05418 0.9082 7
Model 1 Auditory	0.46261 0.2959 7	0.35355 0.4366 7	0.35355 0.4366 7
Model 1 Kinesthetic	0.77030 0.0427 7	0.52931 0.2218 7	0.52931 0.2218 7
Model 1 Mean Workload	0.99152 0.0001 7	0.65886 0.1075 7	0.65886 0.1075 7
Model 1 Overload Conditions	0.83156 0.0204 7	0.87727 0.0095 7	0.87727 0.0095 7
Model 1 Overload Density	0.83594 0.0191 7	0.85562 0.0141 7	0.85562 0.0141 7

	Model 2 Mean Workload	Model 2 Overload Conditions	Model 2 Overload Density
Model 2 Cognitive	0.87438 0.0100 7	0.63969 0.1218 7	0.63969 0.1218 7
Model 2 Psychomotor	0.80236 0.0299 7	0.56353 0.1877 7	0.56353 0.1877 7
Model 2 Visual	0.47071 0.2864 7	0.14859 0.7505 7	0.14859 0.7505 7
Model 2 Auditory	0.46261 0.2959 7	0.35355 0.4366 7	0.35355 0.4366 7
Model 2 Mean Workload	1.00000 0.0 7	0.67241 0.0979 7	0.67241 0.0979 7
Model 2 Overload Conditions	0.67241 0.0979 7	1.00000 0.0 7	1.00000 0.0 7
Model 2 Overload Density	0.67241 0.0979 7	1.00000 0.0 7	1.00000 0.0 7